

# STIC Search Report

Biotech-Chem Library



STIC Database Tracking Number: 125783

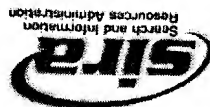
TO: Minh-Tam Davis  
Location: REM-3A24&3C18  
Art Unit: 1642  
Tuesday, June 29, 2004  
Case Serial Number: 10/071521

From: Barb O'Bryen  
Location: Biotech-Chem Library  
Remsen 1A69  
Phone: 571-272-2518  
barbara.obryen@uspto.gov

Search Notes

# RUSH

7/2201



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 18:01:51 ; Search time 127 Seconds  
(without alignments)  
404.956 Million cell updates/sec

Title: AAC50492  
Perfect score: 852  
Sequence: 1 MADEKLPPGWKMRSSG.....GEMSGPVFTDSCIHLIRTE 163

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25.\*

1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriaph.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match		Length	DB ID	Description
		%				
1	753	88.4	159	13	Q919K6	Q919K6 xenopus lae
2	436	51.2	168	5	Q9N492	Q9N492 caenorhabdi
3	397	46.6	176	3	Q42735	Q42735 emericella
4	369.5	43.4	177	3	Q9C475	Q9C475 candida alb
5	364	42.7	178	3	Q8N0N5	Q8N0N5 cryptococcu
6	351.5	41.3	243	5	P90527	P90527 dictyosteli
7	301	35.3	119	10	Q9SL42	Q9SL42 arabidopsis
8	296.5	34.8	121	10	Q94G00	Q94G00 malus domes
9	283	33.2	118	10	Q9LEK8	Q9LEK8 digitalis l
10	280	32.9	123	10	Q7XTJ8	Q7XTJ8 cryza sativ
11	277	32.5	386	5	Q81R35	Q81R35 drosophila
12	219	25.7	69	10	Q42334	Q42334 arabidopsis
13	218.5	25.6	150	5	Q8SR55	Q8SR55 encephalito
14	185	21.7	630	16	Q82SU8	Q82SU8 nitrosomona
15	181	21.2	246	16	Q899G4	Q899G4 clostridium
16	179.5	21.1	92	1	O74049	O74049 cenarchaeum

17	177	20.8	247	16	Q97MB9	Q97MB9 clostridium
18	158	18.5	142	10	Q9FEI8	Q9FEI8 arabidopsis
19	157.5	18.5	126	5	Q9NAF9	Q9NAF9 caenorhabdi
20	157.5	18.5	130	5	Q9VBU4	Q9VBU4 drosophila
21	153.5	18.0	331	16	Q8YUG0	Q8YUG0 bruceella me
22	153.5	18.0	331	16	Q8FYE0	Q8FYE0 xylella fas
23	152	17.8	655	16	Q87E47	Q87E47 xylella fas
24	151	17.7	299	16	Q8CXK4	Q8CXK4 xanthomonas
25	150	17.6	93	16	Q8PBY3	Q8PBY3 rhodospirillum rubrum
26	150	17.6	656	16	Q8PBY3	Q8PBY3 rhodospirillum rubrum
27	149.5	17.5	248	16	Q8XNR4	Q8XNR4 xanthomonas
28	149.5	17.5	280	16	Q81CB1	Q81CB1 clostridium
29	148.5	17.4	286	16	Q81GY5	Q81GY5 bacillus ce
30	148.5	17.4	333	16	Q9KDN4	Q9KDN4 bacillus ha
31	147.5	17.3	287	16	Q81U45	Q81U45 bacillus an
32	147	17.3	655	16	Q9PE37	Q9PE37 xylella fas
33	146	17.1	656	16	Q8PNI2	Q8PNI2 xanthomonas
34	145	17.0	347	16	Q9JSP0	Q9JSP0 neisseria m
35	145	17.0	348	16	Q9KJ86	Q9KJ86 neisseria m
36	143.5	16.8	248	2	Q8GCH6	Q8GCH6 bacillus ce
37	143.5	16.8	283	16	Q81DT1	Q81DT1 bacillus ce
38	142.5	16.7	434	2	Q7X300	Q7X300 uncultured
39	140.5	16.5	298	16	Q81QT1	Q81QT1 bacillus an
40	138.5	16.3	621	16	Q912T8	Q912T8 pseudomonas
41	136.5	16.0	306	16	Q8R760	Q8R760 thermomane
42	136.5	16.0	325	16	Q8CNR4	Q8CNR4 staphylococ
43	134.5	15.8	448	16	Q82W17	Q82W17 nitrosomona
44	134	15.7	93	16	Q821A0	Q821A0 shigella fl
45	133.5	15.7	303	16	Q98G68	Q98G68 rhizobium l

## ALIGNMENTS

RESULT 1  
Q919K6  
ID Q919K6 PRELIMINARY; PRT; 159 AA.  
AC Q919K6  
DT 01-OCT-2000 (TREMREL. 15, Created)  
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)  
DE 01-JUN-2003 (TREMREL. 24, Last annotation update)  
DE Prolyl isomerase P1n1.  
GN P1N1.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20155035; PubMed=10698738;  
RA Winkler K.E., Swenson K.I., Kornbluth S., Means A.R.;  
RT Requirement of the prolyl isomerase p1n1 for the replication  
RT checkpoint.  
RL Science 287:1644-1647(2000).  
DR EMBL; AF239760; AAF43897.1; -.  
DR HSSP; Q13526; P1N1  
DR GO; GO:0016853; P:isomerase activity; IEA.  
DR InterPro; IPR000297; Rotamase.  
DR InterPro; IPR002349; WW.  
DR Pfam; PF00639; Rotamase; 1.  
DR Pfam; PF00397; WW; 1.  
DR PRINTS; PR00403; WWDOMAIN.  
DR SMART; SM00456; WW; 1.  
DR PROSITE; PS01096; PPIC PPIASE\_1; 1.  
DR PROSITE; PS01098; PPIC PPIASE\_2; 1.  
DR PROSITE; PS01159; WW DOMAIN\_1; 1.  
DR PROSITE; PS00020; WW DOMAIN\_2; 1.  
KW Isomerase.  
SQ SEQUENCE 159 AA; 17671 MW; 4E54F66F16C9DF1C CRC64;  
Query Match 88.4%; Score 753; DB 13; Length 159;

```

Db      59 HDGRNPSPSRSDHITRSKDDAINILKSDALNPFHYEKEKLKASNTIEGKREFLAKQFSDC 118
QY      114 SSAKARGDILGAPSRGOMQKPFEDASFALRTGEMSGPVFTDSSGHIILR 161
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      119 SSAKRGDILGPFERRQMOKPFEDASFALRTGEMSDIVDTSSGVHLIYR 166
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
O42735 PRELIMINARY; PRT; 176 AA.
ID O42735
AC AC
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Peptidyl-prolyl cis/trans isomerase.
GN PINA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OC NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R153;
RX MEDLINE=98151356; PubMed=9482729;
RT Crenshaw D.G., Yang J., Means A.R., Kornbluth S.;
RT "The mitotic peptidyl-prolyl isomerase, Pin1, interacts with Cdc25 and
RT Plx1.";
RL EMOB J. 17:1315-1327(1998).
RL EMBL; AF035768; AAC49984.1; -.
DR HSSP; Q13526; IPIN.
DR GO; GO:0016853; F:isomerase activity; IEA.
DR InterPro; IPR000297; Rotamase.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00639; Rotamase; 1_WWP.
DR Pfam; PF00397; WW; 1.
DR SMART; SM00456; WW; 1.
DR PROSITE; P550198; PPIC_PPIASE_2; 1.
DR PROSITE; P501159; WW_DOMAIN_1; 1.
DR PROSITE; P550020; WW_DOMAIN_2; 1.
DR Isomerase.
KW SEQUENCE 176 AA; 20038 MW; 1D1E8376239E1309 CRC64;

Query Match 46.6%; Score 397; DB 3; Length 176;
Best Local Similarity 48.0%; Pred.No.2.6e-29;
Matches 82; Conservative 24; Mismatches 51; Indels 14; Gaps 2;

QY      7 LPFGMEKMSRSRSGRVYFNHITNASQWERPSGNS-----SSGKNGQGQBP-- 53
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      6 LEAGWEVRHSNKNLPYFNPATRESRWEPPADTMEILKMYATVHSGAATYHEAPSQE 65
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      54 -RVRCSHLLVKKHSQRRRSPSWQEKITRTKEALELINGYIOKIKSGEDPFSLASQFSD 112
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      66 GKIRCSHLLVKKHDSRRPSSWREAITRTKEAREILRGHQRIWRGEIRLGLDLAMSESD 125
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY      113 CSSAKARGDLGAFSGOMQKPFEDASFALRTGEMSGPVFTDSSGHIILRTE 163
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      126 CSSAKRGDLGFFGEGMOKFEFAAFALQCVSDIVESGGLHIERLQ 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
Q9C475 PRELIMINARY; PRT; 177 AA.
ID Q9C475
AC AC
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Peptidyl prolyl cis/trans isomerase.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OC NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Devasahayam G., Chaturvedi V., Hanes S.D.;
RT "Isolation of the Candida albicans homolog of the Ess1 gene, encoding
RL a prolyl isomerase.", to the EMBL/GenBank/DBJ databases.
DR SubMITTED (JAN-2000)
DR EMBL: AF224269; AAK00626.1; -
DR HSSP: Q13526; IPIN.
DR GO: GO:0016853; F: isomerase activity; IEA.
DR InterPro: IPR000297; Rotamase.
DR InterPro: IPR001202; WW_Rsp5_WWP.
DR Pfam: PF00639; Rotamase; 1.
DR Pfam: PF00397; WW; 1.
DR SMART: SM00456; WW; 1.
DR PROSITE: PS01198; PPIC_PPIASE_2; 1.
DR PROSITE: PS01159; WW_DOMAIN_1; 1.
DR PROSITE: PS50020; WW_DOMAIN_2; 1.
DR KX Isomerase.
SQ SEQUENCE 177 AA; 19855 MW; 6AE035D939E778E CRC64;

Query Match 43.4%; Score 369.5; DB 3; Length 177;
Best Local Similarity 43.2%; Pred. No. 9.8e-27;
Matches 73; Conservative 34; Mismatches 49; Indels 13; Gaps 2;

QY 7 LPQWEXKMSRSGRVYFHHITNASOWERPSGN-----SSSGKNGQGEPPAR 54
Db 8 LPNFWTIRVSRSHKKEYFLNQSTNESSWDPPYGTDEVLNAYIAKFKNGYKPLVNEGQ 67

QY 55 VRCSHLVKSQRRPSSNQ-EKITRKEALELINGYQIKKSGEDFESLASQFSDC 113
Db 68 VRYSHLLIKNNQKPKXSWSPGISTRDESQILKXHLERILSGEVKULSANTESDC 127

QY 114 SSARKAGDLGAFSGQKQKPFEDASFAIRTGEMSGPVFTDSGHIILRT 162
Db 128 SSHDRGDLGFFSGQKQPPFEAAFNLVHGEVSNIIETNSGVHILQRT 176

RESULT 5
QBUN5 PRELIMINARY; PRT; 178 AA.
AC QBUN5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Prolyl isomerase Ess1.
OS Cryptococcus neoformans var. neoformans.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX NCBI_TaxID=40410;
RN [1]
RP SEQUENCE FROM N.A.
RA Ren P., Chaturvedi V., Hanes S.D.;
RT "The Ess1 prolyl isomerase is required for virulence in Cryptococcus
RT neoformans.", to the EMBL/GenBank/DBJ databases.
RL Submitted (JUL-2002)
DR EMBL: AF533511; AAN03477.1; -
DR GO: GO:0016853; F: isomerase activity; IEA.
DR InterPro: IPR000297; Rotamase.
DR InterPro: IPR001202; WW_Rsp5_WWP.
DR Pfam: PF00639; Rotamase; 1.
DR Pfam: PF00397; WW; 1.
DR SMART: SM00456; WW; 1.
DR PROSITE: PS01198; PPIC_PPIASE_2; 1.
DR PROSITE: PS01159; WW_DOMAIN_1; 1.
DR PROSITE: PS50020; WW_DOMAIN_2; 1.
DR KX Isomerase.
SQ SEQUENCE 178 AA; 19529 MW; 6AE3P317E014784B CRC64;

Query Match 42.7%; Score 364; DB 3; Length 178;
Best Local Similarity 45.5%; Pred. No. 3.2e-26;
Matches 81; Conservative 22; Mismatches 45; Indels 30; Gaps 4;

QY 10 GWEKRMSSRSGRVYFHHITNASOWERPSGNSS-----SSGKNGQ 49
Db 5 GWEIRFSNRQIPFYNSERSISTWEPSELSAEQIQQLFGAAKVMNVQLAPAGGKQ 64

```

```

QY 50 GEPARVRCSHLVKHSQRRPSSNQKQKTRTKEALELIN---GYIQKIKSGE--EDFE 104
Db 65 -----VRASHILAKHAGSRPASPARNVIRITTSDEAQAIIEQHIAYLQSLPPADLPKEFA 119
QY 105 SLASQFSDCSSAKARGDLGAFSGQKQKPFEDASFAIRTGEMSGPVFTDSGHIILRT 162
Db 120 KIASTESDCSSARKGDLGFWGRGQKQKPFEDATFNTPVGQLSGIVKTDGHIILRT 177

RESULT 6
P90527 PRELIMINARY; PRT; 243 AA.
AC P90527;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pina (Fragment).
GN PINA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RA Loomis W.F., Iranfar N.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U78757; AAB36960.1; -
DR HSSP: Q13526; IPIN.
DR InterPro: IPR000253; FHA.
DR InterPro: IPR000297; Rotamase.
DR InterPro: IPR008984; SMAD_FHA.
DR Pfam: PF00498; FHA; 1.
DR Pfam: PF00639; Rotamase; 1.
DR SMART: SM00240; FHA; 1.
DR PROSITE: PS50006; FHA_DOMAIN; 1.
DR PROSITE: PS50198; PPI_C_PPIASE_2; 1.
DR NON TER 1
SQ SEQUENCE 243 AA; 26794 MW; 1596001BB784500A CRC64;

Query Match 41.3%; Score 351.5; DB 5; Length 243;
Best Local Similarity 52.8%; Pred. No. 7e-25;
Matches 76; Conservative 15; Mismatches 48; Indels 5; Gaps 2;

QY 18 SSGRVYFHHITNASOWERPSGNSSCGKNGQGEPPARVRCSHLVKHSQRRPSSWQK 77
Db 103 SSKXHFILKGTNTN---PS-SSSSSSSSSEPKVTTCRHLVXHQSRNPSSWRESK 157

QY 78 ITRTKEALELINGYQIKKSGEDFESLASQFSDCSSAKARGDLGAFSGQKQKPFEDA 137
Db 158 ITRTKEARAKINEYRATIIISGATFEDLHKNSDCSSAKRGYLDFFKRGQMRPFEDC 217

QY 138 SPALRTGEMSGPVFTDSGHIILR 161
Db 218 AFSLVKGEVSGIVDTDSGVHILR 241

RESULT 7
Q9SL42 PRELIMINARY; PRT; 119 AA.
AC Q9SL42;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative peptidyl-prolyl cis-trans isomerase
DE (AT2G18040/T27K22.9).
GN AT2G18040 OR T27K22.9/AT2G18040.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]

```





OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 OC lamids; Lamiales; Plantaginaceae; Digitalidae; Digitalis.  
 OX NCBI\_TaxID=49450;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EHRH;  
 RX MEDLINE=21216637; PubMed=11118437;  
 RA Metzner M., Stollner G., Ruecknagel P., Lu K.P., Fischer G.,  
 RA "Functional replacement of the essential esel in yeast by the plant  
 RT parulin D1Par13.";  
 RL J. Biol. Chem. 276:13524-13529(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EHRH;  
 RA Metzner M.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ133755; CAB94994.1; -;  
 DR HSP; Q13526; lPIN.  
 DR GO; GO:0016853; F-isomerase activity; IEA.  
 DR InterPro; IPR000297; Rotamase.  
 DR Pfam; PF00639; Rotamase; 1.  
 DR PROSITE; PS01096; PPIC\_PP1ASE\_1; 1.  
 DR PROSITE; PS01098; PPIC\_PP1ASE\_2; 1.  
 KW Isomerase.  
 SQ SEQUENCE 118 AA; 12834 MW; 131B74FB4AC3F229 CRC64;  
 Query Match 33.2%; Score 283; DB 10; Length 118;  
 Best Local Similarity 51.3%; Pred. No. 7.5e-19;  
 Matches 58; Conservative 17; Mismatches 34; Indels 4; Gaps 1;  
 QY 54 RVRCSSHLVKHSQSRPSSWRQEK-----ITRTKEALELINGYIOKIKSGEEDFESLQ 109  
 Db 5 KVRASHILIKHQSRKRSKWKDPDGLSATRTDDAVSQLSQRLSDPSPSLASR 64  
 QY 110 FSDCSSAKARGDLGAFSRGQMKPFEDASFALRTGEMSGPVFTDGIHILRT 162  
 Db 65 HSHCSSAKRGDGLGPRGQMKPFEEATFALKVGISDIVDTSVGHILKRT 117  
 RESULT 10  
 QYXJ8 PRELIMINARY; PRT; 123 AA.  
 AC QYXJ8 (T-EMBLrel. 25, Created)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)  
 DE OSJNB0020P07.6 protein.  
 GN OSJNB0020P07.6.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,  
 RA Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,  
 RA Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,  
 RA Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,  
 RA Huang Y.C., Li Y., Zhu J.Q., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,  
 RA Chen L., Fan D.L., Zhang J.Q., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,  
 RA Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,  
 RA Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,  
 RA Zhang R.Q., Guan J.P., Hong G.F.,  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL606450; CAB01289.1; -;  
 SQ SEQUENCE 123 AA; 13340 MW; 511DABB2D4935E77 CRC64;  
 Query Match 32.9%; Score 280; DB 10; Length 123;  
 Best Local Similarity 50.9%; Pred. No. 1.5e-18;  
 Matches 59; Conservative 15; Mismatches 38; Indels 4; Gaps 1;

QY 51 EPARVCSHLVKHSQSRPSSWRQEK-----ITRTKEALELINGYIOKIKSGEEDFESL 106  
 Db 7 ETEIVRALHILIKHSGSRKSKWKDPDGRVLSATRTADAAALDLRLQILSGRANFADL 66  
 QY 107 ASQFSDCSSAKARGDLGAFSRGQMKPFEDASFALRTGEMSGPVFTDGIHILRT 162  
 Db 67 AARHSDCSSARRGGDLGTGRRQMKPFEDATFALKVGEMSDTVDTSVGHILRT 122  
 RESULT 11  
 QYXJ5 PRELIMINARY; PRT; 386 AA.  
 AC QYXJ5 (T-EMBLrel. 23, Created)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE CG32845-PA.  
 GN CG32845.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fogle A., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spiers R., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galie R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases	
EMBL; F13919; CAA23077.1; -. HSP; Q13526; lPIN. GO; GO:0016853; Filisomerase activity; IEA. InterPro; IPR000297; Rotamase. Pfam; PF00639; Rotamase; 1. PROSITE; PS01096; PPIC_PPIASE_1; 1. PROSITE; PS01098; PPIC_PPIASE_2; 1. Isomerase.	
NON TER SEQUENCE	1 69 AA; 7333 MW; 03C2DF735ED6F253 CRC64;
Query Match	25.7%; Score 219; DB 10; Length 69;
Best Local Similarity	62.7%; Pred.No. 3.9e-13;
Matches	42; Conservative 13; Mismatches 12; Indels 0; Gaps 0;
QY	96 IKSGEEDFESLASQFCSCSSAKARGDLGAFSRGMQKQKFEDASALRTGEMSGPVFTDSG 155
Ddb	2 IVSGKANFEVATRVSDCSSAKRGDGLGSFGRGOMQKPFEEATVALKVGDISDVITDSG 61
QY	156 IHIILRT 162
Ddb	62 VHIKRI 68
RESULT 13	
Q8RSR5	PRELIMINARY; PRT; 150 AA.
ID	Q8RSR5
AC	Q8RSR5;
DT	01-JUN-2002 (TrEMBLrel. 21, Created)
DD	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE	01-OCT-2002 (TrEMBLrel. 22, Last annotation update) Peptidyl prolyl cis trans isomerase (NIMA-interacting).
GN	ECT06.0330.
OS	Encephalitozoon cuniculi.
OC	Eukaryota; Fungi; Microsporidia; Unikaryoniidae; Encephalitozoon.
OX	NCBI_TaxID=6035;
EN	[1]
RX	SEQUENCE FROM N.A.
RP	STRAIN=GB-M1;
RA	Genoscope;
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RM	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=GB-M1;
RX	MEDLINE=21576510; PubMed=11719806;
RA	Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomazet F., Prenier G., Barbe V., Peyretailade E., Brottier P., Wincker P., Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M., Weissenbach J., Vivares C.F.; "genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi"; Nature 414:450-453(2001).
RL	EMBL; AL590446; CAD25393.1; -. InterPro; IPR000297; Rotamase.
DR	Pfam; PF00639; Rotamase; 1.
DR	PROSITE; PS01098; PPIC_PPIASE_2; 1.
SQ	SEQUENCE 150 AA; 17203 MW; 51AA3C8864421487 CRC64;
Query Match	25.6%; Score 218.5; DB 5; Length 150;
Best Local Similarity	35.0%; Pred.No. 1.1e-12;
Matches	56; Conservative 24; Mismatches 53; Indels 27; Gaps 5;
QY	8 PGWEKRMSRGVRYTFNHITNASQWERSGNSGGKNKGQEPARVCSHLLVKHSQS 67
Ddb	11 PEMWIKLDKDKTGSPFYNTETAERTKRN-----EG----FRYLHLIKHEKS 56
QY	68 REPSWRQEKTIRTKAEALEINGIQ--KIKGEEDFESL-----ASFQSDCSSAKARGD 121
Ddb	57 RKP-----VDMSIDEAFSRIKAIHEDLRAGADKNRFELFKEAALKHSQCSSAKRGD 109
QY	122 LGAFSRGMQKQFPEDASALRTGEMSGPVFTDSGIHILR 161

Db 110 LGFVCGNMKEPEKPAFSLGRGEMSGPVSTPSGFHIIYR 149

## RESULT 14

Q82S08 PRELIMINARY; PRT; 630 AA.  
AC Q82S08;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE PiC-type peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8).  
GN PFID OR NE2206.  
OS Nitrosomonas europaea.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;  
OC Nitrosomonadaceae; Nitrosomonas.  
OX NCBI\_TaxID=915;  
RN [1]  
RP SEQUENCE FROM N.A. / IFO 14298;  
RC STRAIN=ATCC 19718; PubMed=12700255;  
RX MEDLINE=22586410; PubMed=12700255;  
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,  
HAUSER L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,  
RACIERO D.M., HOMMES N.G., WHITTAKER M.M., ARP D.J.;  
RT "Complete genome sequence of the ammonia-oxidizing bacterium and  
obligate chemolithoautotroph Nitrosomonas europaea.";  
RL J. Bacteriol. 185:2759-2773 (2003).  
DR EMBL; BX321863; CAD86118.1; -.  
DR GO; GO:0042027; F:cyclophilin-type peptidyl-prolyl cis-trans . . .; IEA.  
DR GO; GO:0004600; F:cyclophilin; IEA.  
DR GO; GO:0030051; F:FK506-sensitive peptidyl-prolyl cis-trans i. . .; IEA.  
DR GO; GO:0016853; F:isomerase activity; IEA.  
DR InterPro; IPR000297; Rotamase.  
DR Pfam; PF00639; Rotamase; 1.  
DR PROSITE; PS0198; PPIC\_PPIASE\_2; 1.  
KW Isomerase; Complete proteome.  
SQ SEQUENCE 630 AA; 71638 MW; 35F9D1AFFFC34153 CRC64;

Query Match 21.7%; Score 185; DB 16; Length 630;

Best Local Similarity 32.7%; Pred. No. 8.5e-09;  
Matches 50; Conservative 25; Mismatches 50; Indels 28; Gaps 4;

QY 12 EKRMSRSGRVYFNFHITNASQWERPSGSGKNGGEPARVRCSHLLVKHSQSRPS 71  
Db 241 QEQISDEIRKYDEH-----QDEFGQAEERRASHILLS-----VPA 277  
QY 72 SWRQKTRTKEALELINGYQKISGEEDFESLASQFS-DCSSAKARGDLGAFSRGOM 130  
Db 278 DATEQKTSKARAEQI-----LEQVRQDEKLPFLAALSELDPGSAKGGDLGFFARGLM 333  
QY 131 OKPFEDASFALRTGEMSGPVFTDSGHIILRT 163  
Db 334 VKPFDEVFQMGEGIRGEPVTFPGFHIILRT 366

## RESULT 15

Q899G4 PRELIMINARY; PRT; 246 AA.  
AC Q899G4;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative peptidyl-prolyl cis-trans isomerase.  
GN CTC00215.  
OS Clostridium tetani.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1513;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Massachusetts / E88;  
RX MEDLINE=22457253; PubMed=12552129;  
RA Brueggemann H., Baumer S., Fricke W.F., Wierer A., Liesegang H.,  
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,

RA "The genome sequence of Clostridium tetani, the causative agent of  
tetanus disease";  
RT Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321 (2003).  
RL EMBL; AE015936; AAC34863.1; -.  
DR GO; GO:0016853; F:isomerase activity; IEA.  
DR InterPro; IPR000297; Rotamase.  
DR Pfam; PF00639; Rotamase; 1.  
DR PROSITE; PS01096; PPIC\_PPIASE\_1; 1.  
DR PROSITE; PS0198; PPIC\_PPIASE\_2; 1.  
KW Isomerase; Complete proteome.  
SQ SEQUENCE 246 AA; 28688 MW; C6A29D8AC52640A6 CRC64;

Query Match 21.2%; Score 181; DB 16; Length 246;

Best Local Similarity 38.5%; Pred. No. 6.5e-09;  
Matches 45; Conservative 17; Mismatches 33; Indels 22; Gaps 3;

QY 43 SGGKNGGEPARVRCSHLLVKHSQSRPSRWROKTRTKEALELINGYQKISGEED 102  
Db 103 NNNKNMFIEERVSTKHILVE-----TKEEANIVD-----EIKNG-LS 140  
QY 103 FESTASQFDCSSAKARGDLGAFSRGOMQKPFEDASFALRTGEMSGPVFTDSGHI 159  
Db 141 FEEAAKEYSNCPKSGAGGDLGTFGRWVKEFEAEAFEMKEGTISNPVKTFQGYHII 197

Search completed: June 28, 2004, 18:39:08

Job time : 130 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 28, 2004, 14:35:24 ; Search time 4362.63 Seconds  
(without alignments)  
9875.458 Million cell updates/sec  
Title: U49070  
Perfect score: 994  
Sequence: 1 TGTGGCCAGCACCTCGAGG.....CCCAATTAAACCCAGAACCA 994  
Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 3470272 seqs, 21671516995 residues 6940544  
Total number of hits satisfying chosen parameters:  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.ov.\*  
5: gb.pat.\*  
6: gb.ph.\*  
7: gb.pl.\*  
8: gb.pr.\*  
9: gb.ro.\*  
10: gb.sts.\*  
11: gb.sy.\*  
12: gb.un.\*  
13: gb.vi.\*  
14: gb.ba.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	994	100.0	994	6	AX214524
2	994	100.0	994	6	HSU49070
3	994	100.0	1014	6	AR074009
4	994	100.0	1014	6	AR082120
5	994	100.0	1014	6	AR360786
6	956	96.2	1021	9	BC002899
7	892.6	89.8	3125	6	AX747830
8	892.6	89.8	3125	9	AK092970
9	885.6	89.0	996	9	HSU82382
10	885.6	89.0	153750	9	AL391728
11	567.4	57.1	1656	6	AX834874
12	567.4	57.1	1656	9	AK097558
13	567.4	57.1	138036	9	AC008752
14	567.4	57.1	194624	9	AC008742
15	408.6	41.1	954	10	AB009691
16	407.2	41.0	936	10	BC038254
17	375.2	37.7	138152	10	AL954170
18	346.4	34.8	150017	2	AL773509
19	338	34.0	521	9	HUMZE03B04
20	321.2	32.3	164603	2	AC119466
21	321.2	32.3	223996	2	AC094504
22	319	32.1	529	5	AF239760
23	278.4	28.0	194880	2	AC091329
24	277.2	27.9	852	5	BC059553
25	239.6	24.1	194880	2	AC091929
26	216	21.7	64036	2	AC040921
27	208.4	21.0	439	6	AX305342
28	161.4	16.2	993	3	DMU35140
29	159.8	16.1	17559	10	AB009692
30	158.2	15.9	240584	2	AC073676
31	158.2	15.9	240584	2	AC116063
32	158.2	15.9	241086	2	AC106653
33	158.2	15.9	254452	2	AC118348
34	158.2	15.9	255299	2	AC109676
35	115.6	11.6	65675	2	AC026985
36	115.6	11.6	65675	2	AC026985
37	113.4	11.4	589	11	BV093798
38	108.6	10.9	549	8	NCR6023
39	107.4	10.8	770	8	AY374446
40	105.2	10.6	3415	8	AF350242
41	100.6	10.1	607	11	BV093800
42	99	10.0	66669	3	AF017777
43	98.4	9.9	87089	2	AC013980
44	98.4	9.9	171548	3	AC092494
45	98.4	9.9	173508	3	AC011251

ALIGNMENTS

RESULT 1  
AX214524  
LOCUS AX214524 994 bp DNA linear PAT 06-SEP-2001  
DEFINITION Sequence 3 from Patent WO0159155.  
ACCESSION AX214524  
VERSION AX214524.1 GI:15524567  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 Meritet J.F., Dron,M. and Tovey,M.G.  
AUTHORS Interferon-alpha induced genes  
TITLE Patent: WO 0159155-A 3 16-AUG-2001;  
JOURNAL

Pharma Pacific Pty. Ltd. (AU)  
Location/Qualifiers  
1. .994  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
25. .516  
/notes="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC69362.1"  
/db\_xref="GI:15524568"  
/db\_xref="RENTREMBL:CAC69362"  
/translation="MADEEKLPPGWKEMSRSSGRVYVFNHITNASQWERPSGNSSSG  
GKNGQGPAPRVCRSHLLVKKHSQSRPSSWRQEKIITRKEALELINGYIQIKSGEED  
FESLASQFSDSSAKARGDLGAFSRGQKPFEDASFALTRGEMSGPVFTDSGIHIL  
RTE"

ORIGIN

Query Match 100.0%; Score 994; DB 6; Length 994;  
Best Local Similarity 100.0%; Pred. No. 1.4e-199; Indels 0; Gaps 0;  
Matches 994; Conservative 0; Mismatches 0;

QY 1 TGCTGGCCAGCACCTCGAGGGAAGATGCGGACGAGGAGAGCTGCCGCCCGCTGGGAG 60  
DB 1 TGCTGGCCAGCACCTCGAGGGAAGATGCGGACGAGGAGAGCTGCCGCCCGCTGGGAG 60  
QY 61 AAGCGCATGAGCCGAGCTCAGGCGGAGTGTAATCTCAACACATCACTAACGCCAGC 120  
DB 61 AAGCGCATGAGCCGAGCTCAGGCGGAGTGTAATCTCAACACATCACTAACGCCAGC 120  
QY 121 CAGTGGGAGCGGCCAGCGGCAACAGCAGCAGTGGTGCAAAAACGGCGAGGGGAGCCT 180  
DB 121 CAGTGGGAGCGGCCAGCGGCAACAGCAGCAGTGGTGCAAAAACGGCGAGGGGAGCCT 180  
QY 181 GCCAGGCTCCGCTGCTGCACCTGCTGGTGAAGCAGACGAGTCAAGCGGCGCTCGTCC 240  
DB 181 GCCAGGCTCCGCTGCTGCACCTGCTGGTGAAGCAGACGAGTCAAGCGGCGCTCGTCC 240  
QY 241 TGGCGGAGGAGAGATCACCGGACCAAGGAGGAGGCGCTGGAGCTGATCAAGCGGTAC 300  
DB 241 TGGCGGAGGAGAGATCACCGGACCAAGGAGGAGGCGCTGGAGCTGATCAAGCGGTAC 300  
QY 301 ATCCAGAGATCAAGTCCGGAGAGGAGGACTTTGAGTCTCTGGCTTCAAGTTCAGCGAC 360  
DB 301 ATCCAGAGATCAAGTCCGGAGAGGAGGACTTTGAGTCTCTGGCTTCAAGTTCAGCGAC 360  
QY 361 TGCAGCTCAGCCAGGCGGAGAGCTGGGTGCTTTCAGGAGGTCAGATGAGGAG 420  
DB 361 TGCAGCTCAGCCAGGCGGAGAGAGCTGGGTGCTTTCAGGAGGTCAGATGAGGAG 420  
QY 421 CCATTGAGAGCGCTCGTTGCGTGGCGGAGGAGAGTGGCGGCGCTGTTTACG 480  
DB 421 CCATTGAGAGCGCTCGTTGCGTGGCGGAGGAGAGTGGCGGCGCTGTTTACG 480  
QY 481 GATTCGGGATCCATCATCTCCGCACTGAGTGGAGGAGGAGGAGGAGGAGGAGGAG 540  
DB 481 GATTCGGGATCCATCATCTCCGCACTGAGTGGAGGAGGAGGAGGAGGAGGAGGAG 540  
QY 541 CGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600  
DB 541 CGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600  
QY 601 ACCCCCCCTCTCCGACCGTCAACAGTATTTATTTGTTTCCCAATGGCTGGGAGGGG 660  
DB 601 ACCCCCCCTCTCCGACCGTCAACAGTATTTATTTGTTTCCCAATGGCTGGGAGGGG 660  
QY 661 GCGCTTCCAGATTGGGGGCGCTGGGTCCCACTCCCTGCTCCATCCCAAGTTGGGGTGC 720  
DB 661 GCGCTTCCAGATTGGGGGCGCTGGGTCCCACTCCCTGCTCCATCCCAAGTTGGGGTGC 720  
QY 721 GACCGGAGATTCTCCCTTAAGGAATTGATTCAGCAGGAGGAGGAGGAGGAGGAGGAG 780  
DB 721 GACCGGAGATTCTCCCTTAAGGAATTGATTCAGCAGGAGGAGGAGGAGGAGGAGGAG 780

QY 781 GGGCAGTGTGGTGGGAGGGTGTTCCTCAAGAGAGAGCCCTGGTCAGCAGAGCGGCCCGTG 840  
DB 781 GGGCAGTGTGGTGGGAGGGTGTTCCTCAAGAGAGAGCCCTGGTCAGCAGAGCGGCCCGTG 840  
QY 841 TCCCCCAGGCTGTGGAGGAGCAGACTCGAGGCGCGAATTTCTTCTAGTTAGGCACTCC 900  
DB 841 TCCCCCAGGCTGTGGAGGAGCAGACTCGAGGCGCGAATTTCTTCTAGTTAGGCACTCC 900  
QY 901 TCTGTTCACTCGCAAGGTGAACACTCATGCGGACGATGGCCCTCTGAGCAACTGTG 960  
DB 901 TCTGTTCACTCGCAAGGTGAACACTCATGCGGACGATGGCCCTCTGAGCAACTGTG 960  
QY 961 CAGACCTTTTCACCCCCAATTAAACCCAGAACCA 994  
DB 961 CAGACCTTTTCACCCCCAATTAAACCCAGAACCA 994

RESULT 2  
HSU49070  
LOCUS  
DEFINITION  
Human peptidyl-prolyl isomerase and essential mitotic regulator  
(PIN1) mRNA, complete cds.  
ACCESSION  
U49070  
VERSION  
1  
KEYWORDS  
peptidyl-prolyl isomerase; cell cycle; mitotic regulator; NIMA;  
Eas1; NIMA-interacting protein 1 (Pin1).  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 994)  
AUTHORS  
Lu, K.P., Hanes, S.D. and Hunter, T.  
TITLE  
A human peptidyl-prolyl isomerase essential for regulation of  
mitosis 380 (574), 544-547 (1996)  
JOURNAL  
NATURE  
MEDLINE  
96195064  
PUBMED  
8606777  
REFERENCE  
2 (bases 1 to 994)  
AUTHORS  
Lu, K.P. and Hunter, T.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (12-FEB-1996) Kun Ping Lu, Molecular Biology and Virology  
Laboratory, Salk Institute, 10010 North Torrey Pines Rd., La Jolla,  
CA 92037, USA  
FEATURES  
source  
1. .994  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/cell\_line="HeLa"  
gene  
1. .994  
/gene="PIN1"  
CDS  
25. .516  
/gene="PIN1"  
/note="NIMA-interacting protein 1, essential mitotic  
regulator, essential peptidyl-prolyl isomerase"  
/codon\_start=1  
/product="pin1"  
/protein\_id="AAC50492.1"  
/db\_xref="GI:1332710"  
/translation="MADEEKLPPGWKEMSRSSGRVYVFNHITNASQWERPSGNSSSG  
GKNGQGPAPRVCRSHLLVKKHSQSRPSSWRQEKIITRKEALELINGYIQIKSGEED  
FESLASQFSDSSAKARGDLGAFSRGQKPFEDASFALTRGEMSGPVFTDSGIHIL  
RTE"

ORIGIN

Query Match 100.0%; Score 994; DB 9; Length 994;  
Best Local Similarity 100.0%; Pred. No. 1.4e-199;  
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCTGGCCAGCACCTCGAGGGAAGATGCGGACGAGGAGAGCTGCCGCCCGCTGGGAG 60  
DB 1 TGCTGGCCAGCACCTCGAGGGAAGATGCGGACGAGGAGAGCTGCCGCCCGCTGGGAG 60

[illegible]



```

QY      841  TCCGCCAGGTCGTGGAGCAGACTCGAGGCGCGAATTTGTTTCTAGTTAGGCCACGCTCC 900
Db      841  TCCGCCAGGTCGTGGAGCAGACTCGAGGCGCGAATTTGTTTCTAGTTAGGCCACGCTCC 900
QY      901  TCTGTTAGTCGCAAGGTGAACACTCATGCGGCGACCATGGCCCTCTGAGCAACTGTG 960
Db      901  TCTGTTAGTCGCAAGGTGAACACTCATGCGGCGACCATGGCCCTCTGAGCAACTGTG 960
QY      961  CAGACCCCTTTCACCCCAATTAACCCAGAACCA 994
Db      961  CAGACCCCTTTCACCCCAATTAACCCAGAACCA 994

RESULT 4
LOCUS      AR082120      1014 bp      DNA      linear      PAT 31-AUG-2000
DEFINITION Sequence 1 from patent US 5972697.
ACCESSION AR082120
VERSION    AR082120.1 GI:10008846
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 1014)
AUTHORS    Hunter, T. and Lu, K. Ping.
TITLE      NIMA interacting proteins
JOURNAL    Patent: US 5972697-A 1 26-OCT-1999;
FEATURES   Location/Qualifiers
            source
            1..1014
            /organism="unknown"
            /mol_type="unassigned DNA"

ORIGIN
Query Match      100.0%; Score 994; DB 6; Length 1014;
Best Local Similarity 100.0%; Pred. No. 1.4e-199;
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  TGTGGCCAGCACCTCAGGAGAGAGTGGCGGACGAGGAGAGTGGCGCGCGCTGGGAG 60
Db      1  TGTGGCCAGCACCTCAGGAGAGAGTGGCGGACGAGGAGAGTGGCGCGCGCTGGGAG 60
QY      61  AAGCGCATGAGCCGAGCTCAGGCGGAGTGTACTTCAACACATCACTAAGCCGAGC 120
Db      61  AAGCGCATGAGCCGAGCTCAGGCGGAGTGTACTTCAACACATCACTAAGCCGAGC 120
QY      121  CAGTGGGAGCGCCCGAGCGCAACAGCAGCAGTGGTGGCAAAAACGGCGAGGGAGCCT 180
Db      121  CAGTGGGAGCGCCCGAGCGCAACAGCAGCAGTGGTGGCAAAAACGGCGAGGGAGCCT 180
QY      181  GCCAGGGTCCGCTGCTGCGACCTGCTGGTGAAGCAGCAGCCAGTCAAGCGGCGCTCGTCC 240
Db      181  GCCAGGGTCCGCTGCTGCGACCTGCTGGTGAAGCAGCAGCCAGTCAAGCGGCGCTCGTCC 240
QY      241  TGGCGGCGAGGAGAGATCAACCGGACCAAGGAGAGGCGCTGGAGCTGATCAACGGCTAC 300
Db      241  TGGCGGCGAGGAGAGATCAACCGGACCAAGGAGAGGCGCTGGAGCTGATCAACGGCTAC 300
QY      301  ATCCAGAGATCAAGTCGCGGAGAGAGACTTTGAGTCTCTGGCCCTCAGTTTCAAGCAGC 360
Db      301  ATCCAGAGATCAAGTCGCGGAGAGAGACTTTGAGTCTCTGGCCCTCAGTTTCAAGCAGC 360
QY      361  TGCAGTCAAGCAAGCCAGGCGGAGACCTGGGTCCTTCCAGCAGAGGTTCAGATGAGAGAG 420
Db      361  TGCAGTCAAGCAAGCCAGGCGGAGACCTGGGTCCTTCCAGCAGAGGTTCAGATGAGAGAG 420
QY      421  CCATTTGAAGACGCTTCGTTTGGCTGCGGACGCGGAGAGATGAGCGGCGCCGTGTTTACG 480
Db      421  CCATTTGAAGACGCTTCGTTTGGCTGCGGACGCGGAGAGATGAGCGGCGCCGTGTTTACG 480
QY      481  GATTCGGCATCCACATCAATCCCTCGCACTGAGTGAAGCAGCAGCTGACGGGCGCTCGCTCC 540
Db      481  GATTCGGCATCCACATCAATCCCTCGCACTGAGTGAAGCAGCAGCTGACGGGCGCTCGCTCC 540

```

```

QY      541  CGGGCAGGGCAGGGCGGCTAGGCGGCGCAGCTCCCGCTTCCCGCCAGCCAGCTGGCGGA 600
Db      541  CGGGCAGGGCAGGGCGGCTAGGCGGCGCAGCTCCCGCTTCCCGCCAGCCAGCTGGCGGA 600
QY      601  ACCCGCCACTCCCTGCGCACCGTCAACAGTATTTATTTGTTCCACAAATGGCTGGAGGG 660
Db      601  ACCCGCCACTCCCTGCGCACCGTCAACAGTATTTATTTGTTCCCAATGGCTGGAGGG 660
QY      661  GCCCTTCAGATTGGGGCCCTGGGGTCCCGCTCCCTGTCCATCCCGCAGTTGGGGTGC 720
Db      661  GCCCTTCAGATTGGGGCCCTGGGGTCCCGCTCCCGCTCCATCCCGCAGTTGGGGTGC 720
QY      721  GACCGCCAGATTCTCCCTTAAAGGAATTGACTTCAGCAGGGGTGGAGGCTCCAGACCA 780
Db      721  GACCGCCAGATTCTCCCTTAAAGGAATTGACTTCAGCAGGGGTGGAGGCTCCAGACCA 780
QY      781  GGGCAGTGTGGGAGGGGTGTTCCAAAGAGAGAGCGCTGTCAGCAGAGCCCGCGTG 840
Db      781  GGGCAGTGTGGGAGGGGTGTTCCAAAGAGAGAGCGCTGTCAGCAGAGCCCGCGTG 840
QY      841  TCCCGCCAGGTGCTGAGGCGAGCTCGAGGGCGGAAATTTCTAGTTAGGCCACGCTCC 900
Db      841  TCCCGCCAGGTGCTGAGGCGAGCTCGAGGGCGGAAATTTCTAGTTAGGCCACGCTCC 900
QY      901  TCTGTTAGTCGCAAGGTGAACACTCATGCGGCGACCATGGCCCTCTGAGCAACTGTG 960
Db      901  TCTGTTAGTCGCAAGGTGAACACTCATGCGGCGACCATGGCCCTCTGAGCAACTGTG 960
QY      961  CAGACCCCTTTCACCCCAATTAACCCAGAACCA 994
Db      961  CAGACCCCTTTCACCCCAATTAACCCAGAACCA 994

RESULT 5
LOCUS      AR360786      1014 bp      DNA      linear      PAT 17-AUG-2003
DEFINITION Sequence 1 from patent US 6596848.
ACCESSION AR360786
VERSION    AR360786.1 GI:33768227
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 1014)
AUTHORS    Hunter, T. and Lu, K. P.
TITLE      Antibodies to NIMA interacting proteins
JOURNAL    Patent: US 6596848-A 1 22-JUL-2003;
FEATURES   Location/Qualifiers
            source
            1..1014
            /organism="unknown"
            /mol_type="genomic DNA"

ORIGIN
Query Match      100.0%; Score 994; DB 6; Length 1014;
Best Local Similarity 100.0%; Pred. No. 1.4e-199;
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  TGTGGCCAGCACCTCAGGAGGAGATGGCGGACGAGGAGAGTGGCGCGCGCTGGGAG 60
Db      1  TGTGGCCAGCACCTCAGGAGGAGATGGCGGACGAGGAGAGTGGCGCGCGCTGGGAG 60
QY      61  AAGCGCATGAGCCGAGCTCAGGCGGAGTGTACTTCAACACATCACTAAGCCGAGC 120
Db      61  AAGCGCATGAGCCGAGCTCAGGCGGAGTGTACTTCAACACATCACTAAGCCGAGC 120
QY      121  CAGTGGGAGCGCCCGAGCGCAACAGCAGCAGTGGTGGCAAAAACGGCGAGGGAGCCT 180
Db      121  CAGTGGGAGCGCCCGAGCGCAACAGCAGCAGTGGTGGCAAAAACGGCGAGGGAGCCT 180
QY      181  GCCAGGGTCCGCTGCTGCGACCTGCTGGTGAAGCAGCAGCTGACGGGCGCTCGCTCC 240
Db      181  GCCAGGGTCCGCTGCTGCGACCTGCTGGTGAAGCAGCAGCTGACGGGCGCTCGCTCC 240

```



QY	241	TGCGGCGAGGAGAGATACACCGGACCAAGGAGGCGCTGGAGCTGATCAACGGCTAC	300
Db	241	TGCGGCGAGGAGAGATACACCGGACCAAGGAGGCGCTGGAGCTGATCAACGGCTAC	300
QY	301	ATCCAGAGATCAAGTCGGGAGAGGAGGACTTTGAGTCTCTGCGCTCACAGTTTCAGCGAC	360
Db	301	ATCCAGAGATCAAGTCGGGAGAGGAGGACTTTGAGTCTCTGCGCTCACAGTTTCAGCGAC	360
QY	361	TGCAGCTCAGCCCAAGGCGAGGAGAGACCTGGGTGCTTCAGCGAGGTGATGCGAGAAG	420
Db	361	TGCAGCTCAGCCCAAGGCGAGGAGAGACCTGGGTGCTTCAGCGAGGTGATGCGAGAAG	420
QY	421	CCATTTGAAGACCGCTGTTTGCGTGCAGCGGAGAGATGAGCGGCGCGCTGTTTCACG	480
Db	421	CCATTTGAAGACCGCTGTTTGCGTGCAGCGGAGAGATGAGCGGCGCGCTGTTTCACG	480
QY	481	GATTCGCGCATCCACATCATCTCCGCACTGAGTGAGGGTGGGAGCCCGGCTGCGCT	540
Db	481	GATTCGCGCATCCACATCATCTCCGCACTGAGTGAGGGTGGGAGCCCGGCTGCGCT	540
QY	541	CGGCGCAGGCGAGGCGGCTAGGCGGCGAGCTCCCTTCCCGCCAGCCAGTGGCGGA	600
Db	541	CGGCGCAGGCGAGGCGGCTAGGCGGCGAGCTCCCTTCCCGCCAGCCAGTGGCGGA	600
QY	601	ACCCGCCACTCCCTGCCACCGTCAACAGTATTTATTGTTCCCAATGGCTGGAGGGG	660
Db	601	ACCCGCCACTCCCTGCCACCGTCAACAGTATTTATTGTTCCCAATGGCTGGAGGGG	660
QY	661	GCCCTTCAGATGGGGGCGCTGGGGTCCCGACCTCCCTGTCATCCCGAGTGGGGTGC	720
Db	661	GCCCTTCAGATGGGGGCGCTGGGGTCCCGACCTCCCTGTCATCCCGAGTGGGGTGC	720
QY	721	GACCGCCAGATTCCTCCCTTAAGGAATTCATTCAGCAGGGTGGGAGCTCCGACACCA	780
Db	721	GACCGCCAGATTCCTCCCTTAAGGAATTCATTCAGCAGGGTGGGAGCTCCGACACCA	780
QY	781	GGGAGTGTGGTGGAGGGGTGTTCCAAAGAGAGCGCTGGTCAGCAGCGCGCCCGCTG	840
Db	781	GGGAGTGTGGTGGAGGGGTGTTCCAAAGAGAGCGCTGGTCAGCAGCGCGCCCGCTG	840
QY	841	TCCCCCGAGGTGCTGGAGGAGAGTCCGAGGCGCGAATGTTTCTAGTTAGGCCACGCTCC	900
Db	841	TCCCCCGAGGTGCTGGAGGAGAGTCCGAGGCGCGAATGTTTCTAGTTAGGCCACGCTCC	900
QY	901	TCTGTTTCAGTCCGAAGGTGAACATCATCGCGGAGCCATGGCGCTCTGACCACTGTG	960
Db	901	TCTGTTTCAGTCCGAAGGTGAACATCATCGCGGAGCCATGGCGCTCTGACCACTGTG	960
QY	961	CAGACCTTTACCCCGCAATTAACCCGAAACCA 994	
Db	961	CAGACCTTTACCCCGCAATTAACCCGAAACCA 994	
RESULT	6		
LOCUS	BC002899	1021 bp mRNA linear PRI 12-NOV-2003	
DEFINITION	Homo sapiens protein (peptidyl-prolyl cis/trans isomerase)		
	NIMA-interacting 1, mRNA (CDNA clone MGC:10717 IMAGE:3941595), complete cds.		
ACCESSION	BC002899	1 GI:12804092	
VERSION	MGC.		
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 1021)		
AUTHORS	Srausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shennen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,		

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J., McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, B., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A.G., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A. full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

2 (bases 1 to 1021)  
12477932  
JOURNAL MEDLINE  
REFERENCE Strausberg, R.  
AUTHORS Direct Submission  
TITLE Submitted (05-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
JOURNAL  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Rubin Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@hgrl.nih.gov](mailto:nisc_mgc@hgrl.nih.gov)  
Ahter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, P., Blakesley, R.W., Bouffard, G.G., Breen, K., Gupta, J., Haghighi, P., Dietrich, N.L., Granite, S., Guan, X., Li, J., Liao, J., Liao, J., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, D.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAL plate: 15 Row: m Column: 9  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5453897.

## FEATURES

## source

1..1021  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:10717 IMAGE:3941595"  
/tissue\_type="Lung, small cell carcinoma"  
/clone\_lib="NIH MGC 7"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"

## gene

1..1021  
/gene="PIN1"  
/note="synonym: DOD"  
/db\_xref="LocusID:5300"  
/db\_xref="MIM:601052"  
28..519  
/codon\_start=1  
/product="protein (peptidyl-prolyl cis/trans isomerase)"  
NIMA-interacting 1"

## CDS

/protein\_id="AAH02899.1"  
/db\_xref="GI:12804093"  
/db\_xref="LocusID:5300"  
/translation="MADEKLPGWEKMSRSGRVYFNHITNASQWEPNSGSSG  
GKNGQEPARVRCSHLLVVKHSQSRRTSSRQEKITTKKEALELINGYIQIKSGED"

misc feature

		96.2%; Score 956; DB 9; Length 1021;	
	Best Local Similarity	99.2%; Pred. No. 1.5e-191;	
	Matches	Conservative 0; Mismatches 5; Indels 3; Gaps 2;	
QY	8	CAGCACTCAGAGGAAGATGGCGGACGAGAGAAGCTGCCGCCGGTGTGGAGAAGCGCA	67
DB	11		70
QY	68	TGAGCCGCAGCTCAGGCCGAGTGTAATTCAAACAATCATCTAATCCGACGACGTTGGG	127
DB	71		130
QY	128	AGCGGCCAGGGCCAAACAGCAGCAGTGTCGCAAAAACGGCGACGGGGGAGCCCTGCCAGGG	187
DB	131		190
QY	188	TCCGCTGCTCGCACCTGCTGTAAGACAACGCCAGTCAACGCGGCCCCTGCTCTCGGCGC	247
DB	191		250
QY	248	AGGAAGAATCACCCGACCAAGGAGGAGGCCCTGGAGCTGATCAAACGGCTACATCCAGA	307
DB	251		310
QY	308	AGATCAAGTCGGGAGAGGAGCACTTTGAGTCTCTGGCCTCACAGTTCAGCGACTCGAGCT	367
DB	311		370
QY	368	CAGCCAAGGCCAGGGGAGACCTGGGTGCTTCACGACAGGTCAGATGCAGAAAGCCATTG	427
DB	371		430
QY	428	AAGACGCTCTGTTTCGCTGCGGACGGGGGAGATGAGCGGCCCGGTGTTCAACGATTCCG	487
DB	431		490
QY	488	GCAATCCAATCATCTCCGACATGAGGTGGGGAGCCACGAGCTGGCTCGGGGCA	547
DB	491		550
QY	548	GGGCAAGGCGGCTAGCGCGGCCAGCTCCGCCCTTGCCCGCACGACAGTGGCCACC	607
DB	551		610
QY	608	ACTCCCTGCCACCGTCAACAGATATTATTGTTCCCAATAATGGCTGGGAGGGGGCCCTTC	667
DB	611		670
QY	668	CAGATTGGGGGCCCTGGGTCCCCACTCCCTGTCCATCCCGAGTGGGGCTGCGACCGCC	727
DB	671		730
QY	728	AGATTCTCCCTTAAAGAAATTGACTTCAGAGGGGTGGGAGGTCTCCAGACCCAGGGCAGT	787
DB	731		790
QY	788	GTGGTGGGAGGGTGTTCCAAAGAGAAGCCCTGGTTCAGCAGAGCCGCCCTGTCTCCCGCC	847

Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES	source	Location/Qualifiers	1..3125	organism="Homo sapiens"	/mol_type="mRNA"	/db_xref="taxon:9606"	/clone="SPLEN2013670"	/tissue_type="spleen"	/clone_lib="SPLEN2"	/note="cloning vector: pME18SFL3"	ORIGIN
		Query Match	89.8%	Score 892.6; DB 9; Length 3125;							
		Best Local Similarity	98.7%	Pred. No. 3.3e-178;							
		Matches	921; Conservative	0; Mismatches	9; Indels	3; Gaps	2;				
QY	65	GCATGAGCCGACGCTCAGGCCGAGTGTACTTCAACACATCATCACTAACCGCCAGCCAGT	124								
Db	2191	GCCTGCCCTCCCTCCAGGCCGAGTGTACTTCAACCCATCATCTAACCGCCAGCCAGT	2250								
QY	125	GGGAGCGGCCGACGCGCAACAGCAGCAGTGTGTGGCAAAAAACGGGCGAGGGGAGCTGCCA	184								
Db	2251	GGGAGCGGCCGACGCGCAACAGCAGCAGTGTGTGGCAAAAAACGGGCGAGGGGAGCTGCCA	2310								
QY	185	GGGTCCGCTGCTCGACCTGCTGGTGAAGCAGCAGCAGTGTGTGGCAAAAAACGGGCGAGGGGAGCTGCCA	244								
Db	2311	GGGTCCGCTGCTCGACCTGCTGGTGAAGCAGCAGCAGTGTGTGGCAAAAAACGGGCGAGGGGAGCTGCCA	2370								
QY	245	GGCAGGAGAGATCACCCGGACCAAGGAGGAGGCGCTGGAGCTGATCAACGGCTACATCC	304								
Db	2371	GGCAGGAGAGATCACCCGGACCAAGGAGGAGGCGCTGGAGCTGATCAACGGCTACATCC	2430								
QY	305	AGAAAGATCAAGTCGGGAGAGGAGGACTTTGAGTCTCTGGCTCTCAGTTCAGTACGAGTGC	364								
Db	2431	AGAAAGATCAAGTCGGGAGAGGAGGACTTTGAGTCTCTGGCTCTCAGTTCAGTACGAGTGC	2490								
QY	365	GCTCAGCCAAAGCCAGGCGGAGACTGGTGTCTTTCAGCAGAGGTTCAGTGCAGAGCCAT	424								
Db	2491	GCTCAGCCAAAGCCAGGCGGAGACTGGTGTCTTTCAGCAGAGGTTCAGTGCAGAGCCAT	2550								
QY	425	TTGAAGAGCGCTCTTTCGGCTGGGAGAGAGGAGGCGGCGGAGAGTGCAGGAGTTC	484								
Db	2551	TTGAAGAGCGCTCTTTCGGCTGGGAGAGAGGAGGCGGCGGAGAGTGCAGGAGTTC	2610								
QY	485	CCGGCATCCACATCATCTCCGCACTGAGTGAAGGTGGGAGCCAGGCGCTGGCTCGGG	544								
Db	2611	CCGGCATCCACATCATCTCCGCACTGAGTGAAGGTGGGAGCCAGGCGCTGGCTCGGG	2670								
QY	545	GCAGGCGAGGGGCGCTAGGCGGCGGAGCTCCCTTTCGGCGGAGCCAGCCAGTGGCCGAAACC	604								
Db	2671	GCAGGCGAGGGGCGCTAGGCGGCGGAGCTCCCTTTCGGCGGAGCCAGCCAGTGGCCGAAACC	2730								
QY	605	CCCACCTCCCTCCGACCGTTCACACAGTATTTATTTGTTCCCAATGCCAGTTGGGCGTGG	664								
Db	2731	CCCACCTCCCTCCGACCGTTCACACAGTATTTATTTGTTCCCAATGCCAGTTGGGCGTGG	2790								
QY	665	TTCCAGATTGGGGGCGCTGGGCTCCCACTCCCTGTCCATCCCGAGTTGGGCGTGGGAGC	724								
Db	2791	TTCCAGATTGGGGGCGCTGGGCTCCCACTCCCTGTCCATCCCGAGTTGGGCGTGGGAGC	2850								
QY	725	GCCAGATTCTCCCTTAAGGAATTTGACTTTCAGCAGGGGTGGGAGGCTCCAGAGCCAGGGC	784								
Db	2851	GCCAGATTCTCCCTTAAGGAATTTGACTTTCAGCAGGGGTGGGAGGCTCCAGAGCCAGGGC	2910								
QY	785	AGTGTGTGGGAGGGGTGTTCCAAAGAGAGGCGCTGTGTCAAGAGCGCGCCCGTGTCC	844								
Db	2911	AGTGTGTGGGAGGGGTGTTCCAAAGAGAGGCGCTGTGTCAAGAGCGCGCCCGTGTCC	2970								
QY	845	CCGAGTGTGGGAGGAGACTCGAGGGCGGAAATTTCTAGTTAGGCGAGGCTCTCTG	904								
Db	2971	CCGAGTGTGGGAGGAGACTCGAGGGCGGAAATTTCTAGTTAGGCGAGGCTCTCTG	3030								
QY	905	TTCACTGCGAAAGTGAACACTCATGCGG--CAGCCATGGGCGCTCTGAGCAACTGTGCA	962								
Db	3031	TTCACTGCGAAAGTGAACACTCATGCGGCGGCGGAGCCATGGGCGCTCTGAGCAACTGTGCA	3090								
QY	963	G-ACCCTTTCACCCCAATTAACCCAGAACCA	994								
Db	3091	GCACCCCTTTCACCCCAATTAACCCAGAACCA	3123								
RESULT 8	AK092970	3125 bp	mRNA	linear	PRI 15-JUL-2002						
LOCUS	Homo sapiens	cdna	FLJ35651	f1s, clone SPLEN2013670.							
DEFINITION	AK092970										
ACCESSION	AK092970.1	GI:21751686									
VERSION	oligo capping; f1s (full insert sequence).										
KEYWORDS	Homo sapiens (human)										
SOURCE	Homo sapiens										
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.										
REFERENCE	1										
AUTHORS	Ishibaashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S., Ono, Y., Hotuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S., Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saio, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.										
	NEDO human cdna sequencing project										
JOURNAL	Unpublished										
REFERENCE	2 (bases 1 to 3125)										
AUTHORS	Isogai, T. and Yamamoto, J.										
TITLE	Direct Submission										
JOURNAL	Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan										
	[E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986]										
COMMENT	NEDO human cdna sequencing project supported by Ministry of Economy, Trade and Industry of Japan; CDNA full insert sequencing: Research Association for Biotechnology (RAB); CDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3' - end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and										

Tue Jun 29 08:37:21 2004

```

QY 905 TTCAAGTGCAGAAAGGTGAACACTCATGCGG--CAGCCATGGGGCCCTCTGAGCAACTGTGCA 962
Db 3031 TTCAAGTGCAGAAAGGTGAACACTCATGCGGCGCCAGCCATGGGCCCTCTGAGCAACTGTGCA 3090
QY 963 G-ACCTTTACACCCCAATTAACCCAGAACCA 994
Db 3091 GCACCCCTTACACCCCAATTAACCCAGAACCA 3123

RESULT 9
HSU82382 996 bp mRNA linear PRI 21-OCT-1997
LOCUS Homo sapiens PIN1 peptidyl-prolyl cis/trans isomerase-like (PIN1L)
DEFINITION mRNA, complete cds.
ACCESSION U82382
VERSION U82382.1 GI:2501784
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 996)
AUTHORS Campbell,H.D., Webb,G.C., Fountain,S. and Young,I.G.
TITLE The human PIN1 peptidyl-prolyl cis/trans isomerase gene maps to
human chromosome 19p13 and the closely related PIN1L gene to 1p31
JOURNAL Genomics 44 (2), 157-162 (1997)
MEDLINE 97446133
PUBMED 9299231
REFERENCE 2 (bases 1 to 996)
AUTHORS Campbell,H.D.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1996) Centre for Molecular Structure and
Function, RSBs, Australian National University, PO Box 475,
Canberra ACT 2601, Australia
FEATURES
source 1..996
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p13"
gene 1..996
/genes="PIN1L"
CDS 36..338
/genes="PIN1L"
/notes="contains WW domain related to PIN1"
/codon_start=1
/product="PIN1 peptidyl-prolyl cis/trans isomerase-like"
/protein_id="AAB01333.1"
/db_xref="GI:2501785"
/translaton="MADEKLPQWEKMSRPSGRGVYFNHITNPSOWERPSGNSSSG
GIWQGEFARVRSHLLVKKVKAALDIAAGNHPDQGGPGADQLHPEDQGRGL"
979..984
polyA_signal /gene="PIN1L"
/evidence=experimental
polyA_site 996
/genes="PIN1L"
/evidence=experimental

ORIGIN
Query Match 69.0%; Score 685.6; DB 9; Length 996;
Best Local Similarity 86.1%; Pred. No. 1.e-134;
Matches 861; Conservative 0; Mismatches 84; Indels 55; Gaps 7;

QY 17 GAGGGAAGATGCGGACGAGAGAGCTGCGCCCGCTGGGAGAGCGCATGAGCCGCA 76
Db 28 GAGGAAGATGCGGACGAGAGAACTCGCCCTGGTGGGAAAGCGCATGAGCCGCC 87
QY 77 GCTAGGCGGAGTGTACTTCAACCATCTCAACGCGCAGCGAGTGGAGGGCGGCCA 136
Db 88 CCTCAGGCGGAGGTACTTCAACCATCTCAACGCGCAGCGAGTGGAGGGCGGCCA 147
QY 137 GCGGCAACAGCAGCAGTGTGTGGCAAAACCGGCGAGGGGAGCGCTCCGCGGTCT 196

```

```

Db 148 GCGGCAACAGCAGCAGTGTGTGGCAAAATCTGGAGGGGAGCGCCGAGGTCGCGCGCT 207
QY 197 CGACACTGTGTGAAGCAGCAGCAGTGTGTGGGCGGCGCTCTGCTGTGGCGGAGAGAGA 256
Db 208 CGACACTGTGTGTGAAGCAGCAGCAGTGTGTGGGCGGCGCTCTGCTGTGGCGGAGAG 259
QY 257 TCACCCGAGCAAGAGGAGGAGGCGCTGTGAGTGTGTGAAGTGTGTGAAGTGTGTGAAG 316
Db 260 TCACCCGAGCAAGAGGAGGAGGCGCTGTGAGTGTGTGAAGTGTGTGAAGTGTGTGAAG 319
QY 317 CGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 376
Db 320 CAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 379
QY 377 CCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 436
Db 380 CCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 439
QY 437 CGTTTGGCGCTCGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 496
Db 440 GGTTTGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 499
QY 497 TCATCTCTCGGACACTGTGAGTGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 556
Db 500 TCATCTCTCGGACACTGTGAGTGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 548
QY 557 GGCTAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 616
Db 549 -----TTCCCGCGCTGCTGTGGCGAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 594
QY 617 CACCGCTCAGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 676
Db 595 CACTGTCTCAGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 654
QY 677 GCGCGCT-----GGGTCGCCACTCCCTGTCCTCCATCCCGAGTGGGGCT 718
Db 655 GCGCGCTTCCAGACTGGGGCGCGCTGGGGCTTCCCGACTTCCCTGTGCTGCTGCTGCTGCTGCT 714
QY 719 GCGACCGCGCAGATTTCTCCCTTAAGGAATGACTTTCAGCAGGAGGAGGAGGAGGAGGAGG 778
Db 715 GCGACCTCCAGACTTCTCAATTAAGAGTGTGACTTTCAGCATGGGAGGAGGAGGAGGAGG 774
QY 779 CAGGCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 838
Db 775 CAGGCGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 834
QY 839 TGTGTCGCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 896
Db 835 TGTGTCGCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 894
QY 897 CTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 955
Db 895 CTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 954
QY 956 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 994
Db 955 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 994

```

```

RESULT 10
AL391728 Human DNA sequence from clone RP11-423A18 on chromosome 1, complete
LOCUS sequence.
DEFINITION AL391728
ACCESSION AL391728.19 GI:21217762
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 153750)

```

**AUTHORS**  
**TITLE**  
**JOURNAL**

**COMMENT**

Wallis,J.  
 Direct Submission  
 Submitted (25-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk  
 On May 27, 2002 this sequence version replaced gi:21213126.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em: EMBL; Sw:  
 SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 This sequence was generated from part of bacterial clone contigs of human  
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
 Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr1  
 RP11-423A18 is from the library RP11-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBAC3.6.  
 Location/Qualifiers  
 1. 153750  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /clone="RP11-423A18"  
 /clone\_lib="RP11-11.2"

**FEATURES**  
 source

**ORIGIN**

Query Match 69.0%; Score 685.6; DB 9; Length 153750;  
 Best Local Similarity 86.18; Pred. No. 1.1e-134;  
 Matches 861; Conservative 0; Mismatches 84; Indels 55; Gaps 7;

17 GAGGAAGATGGCGGACGAGGAGAGTGGCGCGCGCTGGGAGAGCGGATGAGCGCA 76  
 Db 107991 GAGGAAGATGGCGGACGAGGAGAACTGGCGCGCTGGGAGAGCGGATGAGCGCGCC 108050

77 GCTCAGGCGGAGTGTACTTCAACCAATCACTAAGCCAGGAGTGGGAGCGGCCA 136  
 Db 108051 CCTCAGGCGGAGGTACTTCAACCAATCACTAAGCCAGGAGTGGGAGCGGCCA 108110

137 GCGCAACAGCAGCAGTGGTGGGCAAAACGGGCGAGGAGCGCTGCCAGGCTCGCGTCT 196  
 Db 108111 GCGGACAGCAGCAGTGGTGGGCAAAATCTGGCAGGAGGAGCGCGCGCTCGCGCT 108170

197 GCGACCTGCTGGTGAAGCAGCAGGAGTCAAGCGGCGCGCTCGTCTGGCGGAGGAGA 256  
 Db 108171 GCGACCTGCTAGTGA-----AGCCAGTCAAGCGGCGCGCTCGACCTGGCGGAGG---AAA 108222

257 TCACCGGACCAAGAGGAGCGCGCTGGAGCTGATCAAGCGGTACATCCAGAGATCAAGT 316  
 Db 108223 TCACCGGACCAAGAGGAGCGCGCTGGAGCTGATCAGCGGCTACATCCAGAGATCAAGG 108282

317 CGGAGAGGAGGACTTTCAGTCTCTGGGCTCACAGTTCAGCGACTGCGAGCTCAGCCAAAG 376  
 Db 108283 CAGGAGAGAGGACTTTCAGTCTCTGGGCTCACAGTTCAGCGACTGCGAGCTCGGCCAAGG 108342

377 CCAGGGGAGACTGGGTGCTTTCAGCAGAGGTCAGATGAGGAGGATTTGAAGACGGCT 436  
 Db 108343 CCAGGGGAGACTGGGTGCTTTCAGCAGAGGTCAGATGAGGAGGATTTGAAGACGGCT 108402

437 CGTTTGGCTCGGCACGGGGGAGATGAGCGGGCCCGTGTTCACGATTCGGGATTCGACAA 496  
 Db 108403 GGTTCGGCGGCGAGCGGGGAGATGAGCGGGAGAGTGTTCACGATTCGGGATTCGACAA 108462

497 TCATCTCCGACATGAGTGAAGGTGGGAGAGCGGAGCGCTGAGCTCGGGGAGGAGGCG 556  
 Db 108463 TCATTCGCGCAGCGAGTGAAGTGGGGGCGCCAGAGCGCTGAGCTCGGGG----- 108511

557 GGCTAGGCGGGCGAGCTCCCGCTTCCCGCGCAGCAGTGGCGGAACCCCGACCTCCCTGC 616  
 Db 108512 -----TTCCCGCGCTGCTGTCGCCAGTGGCGGAACCCCGACCTCCCTGC 108557

617 CACCGCTCACAGATATTTATTTCCCAATGGCTGGAGGGGGCCCTTCCAGATTGGG 676  
 Db 108558 CACTCTCACAGATATTTATTTACCAAAATGGCTGGAGGGGGCCCTTCCAGACTGGG 108617

677 GGCCTT-----GGGTCCCACTCCCTGTCATCCCGAGTGGGGCT 718  
 Db 108618 GGCCTTCCAGACTGGGGGCCCTGGGGTTCCCACTCCCGAGTGGGGCT 108677

719 GCGACCGCGAGATTTCTCCCTTAAGGAATGACTTCAGCAGGGGTGGGAGGCTCCCGAG 778  
 Db 108678 GCGACTCCAGACTTCCCATTAAGGAGTTGACTTCAGCATGGAGGAGGCTCCCGAGAC 108737

779 CAGGGCAGTGTGGTGGGAGGGTGTCCAAAGAGAGGCGCTGGTCAGCAGAGCGGCCCG 838  
 Db 108738 CAGGGCAATGTGGTGGGAGGAGTGTCCAAAGAGAGAGTCTGGTCAGCAGAGCGGCCCG 108797

839 TGTCCCGCCAGGTGTGGAGGAGCAGACTCGAGGGCC--GAATTGTTTCTAGTTAGGCG 896  
 Db 108798 TGTCCCGCCAGGTCTCTGGAGGAGATGCGAGGGCCCGAGTGTGTTGTTAGTTAGGCG 108857

897 CTCTCTGTTCTAGTCGCAAGGTGAACACTCTATGCG--GCAGCCATGGGCCCTCTGAGCA 955  
 Db 108858 CTCTCTGTTCTAGTACAAAGAGCAAACTCTGTCGGCCAGCTGTGGGCCCTCTGAGCA 108917

956 CTGTGCGAG-ACCCTTTTACCCCCCAATTAACCCAGAACCA 994  
 Db 108918 CTGTGCGAGCACCTTTTACCCCCCAATTAATCGGAACCA 108957

**RESULT 11**  
 AX834874  
 LOCUS AX834874 1656 bp DNA linear PAT 15-DEC-2003  
 DEFINITION Sequence 1998 from Patent EP1347046.  
 ACCESSION AX834874  
 VERSION AX834874.1 GI:39921009  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,  
 Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,  
 Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and  
 Masuho, Y.  
 Full-length cDNA sequences  
 Patent: EP 1347046-A 1998 24-SEP-2003;  
 Research Association for Biotechnology (JP)  
 Location/Qualifiers  
 source  
 1. 1656  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"

**FEATURES**  
 source

**ORIGIN**

Query Match 57.1%; Score 567.4; DB 6; Length 1656;  
 Best Local Similarity 98.5%; Pred. No. 1.6e-109;  
 Matches 594; Conservative 0; Mismatches 6; Indels 3; Gaps 2;

395 CCTTCAGCAGAGGTTCAGATGTCAGAGGCCATTTGAAGACGGCTCGTTGCGTTCGCGAGG 454

Db 1045 CCTGGCTCCAGGTCAGATGAGAGCATTGTAAGACGGCTCGTTTGGCTGGGACGG 1104  
QY 455 GGGAGATGAGCGGCGCGTGTTCAGGATTCGGGATCCGACATCATCTCCGCACTGAGT 514  
Db 1105 GGGAGATGAGCGGCGCGTGTTCAGGATTCGGGATCCGACATCATCTCCGCACTGAGT 1164  
QY 515 GAGGTTGGGAGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 574  
Db 1165 GAGGTTGGGAGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1224  
QY 575 CCCCTTGGCGGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 634  
Db 1225 CCCCTTGGCGGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1284  
QY 635 ATTGTTCCCAATGCTGGGAGGCGGCTTCCAGATTTGGGCGGCTGAGCGGCGGCTC 694  
Db 1285 ATTGTTCCCAATGCTGGGAGGCGGCTTCCAGATTTGGGCGGCTGAGCGGCGGCTC 1344  
QY 695 CCCCTTGGCGGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 754  
Db 1345 CCCCTTGGCGGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1404  
QY 755 GCAGGTTGGGAGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 814  
Db 1405 GCAGGTTGGGAGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1464  
QY 815 GCGCTGCTCAGCAGAGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 874  
Db 1465 GCGCTGCTCAGCAGAGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1524  
QY 875 AATTGTTTCTAGTTAGGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 933  
Db 1525 AATTGTTTCTAGTTAGGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1584  
QY 934 -CAGCCATGGGCGGCTGTAGCAACTGTGAGCGGAGCGGCGGCTGAGCGGCGGCTC 991  
Db 1585 -CAGCCATGGGCGGCTGTAGCAACTGTGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1644  
QY 992 CCA 994  
Db 1645 CCA 1647

RESULT 12  
AK097558  
LOCUS  
DEFINITION  
Homo sapiens cDNA FLJ40239 fis, clone TESTI203436, highly similar to PEPTIDYL-PROLYL CIS-TRANS ISOMERASE NIMA-INTERACTING 1 (EC 5.2.1.8).  
ACCESSION  
AK097558  
VERSION  
AK097558.1 GI:21757365  
KEYWORDS  
oligo capping; fis (full insert sequence).  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1  
Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Watanabe,A., Ishii,S., Yamamoto,J., Isono,Y., Matsui,H., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsumoto,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagaue,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Suzuki,Y., Sugano,S., Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 1656)  
REFERENCE  
Isogai,T. and Yamamoto,J.  
Direct Submission  
TITLE  
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
JOURNAL  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT  
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB.  
FEATURES  
Location/Qualifiers  
source  
1. 1656  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="TESTI203436"  
/tissue\_type="testis"  
/clone\_lib="TESTI2"  
/note="cloning vector: pME18SFL3"

ORIGIN  
Query Match 57.1%; Score 567.4; DB 9; Length 1656;  
Best Local Similarity 98.5%; Pred. No. 1.6e-109;  
Matches 594; Conservative 0; Mismatches 6; Indels 3; Gaps 2;  
QY 395 CCTTCAGCAGAGTCAGATGAGAGCGCATTTGAGAGCGCTCGTTTGGCTGGGACGG 454  
Db 1045 CCTGGCTCCAGGTCAGATGAGAGCGCATTTGAGAGCGCTCGTTTGGCTGGGACGG 1104  
QY 455 GGGAGATGAGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 514  
Db 1105 GGGAGATGAGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1164  
QY 515 GAGGTTGGGAGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 574  
Db 1165 GAGGTTGGGAGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1224  
QY 575 CCCCTTGGCGGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 634  
Db 1225 CCCCTTGGCGGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1284  
QY 635 ATTGTTCCCAATGCTGGGAGGCGGCTTCCAGATTTGGGCGGCTGAGCGGCGGCTC 694  
Db 1285 ATTGTTCCCAATGCTGGGAGGCGGCTTCCAGATTTGGGCGGCTGAGCGGCGGCTC 1344  
QY 695 CCTTCATCCCGAGTTGGGCTGCGACCGCGCAGATTTCCCTTAAGGAATTGACTTCA 754  
Db 1345 CCTTCATCCCGAGTTGGGCTGCGACCGCGCAGATTTCCCTTAAGGAATTGACTTCA 1404  
QY 755 GCAGGTTGGGAGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 814  
Db 1405 GCAGGTTGGGAGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1464  
QY 815 GCGCTGCTCAGCAGAGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 874  
Db 1465 GCGCTGCTCAGCAGAGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1524  
QY 875 AATTGTTTCTAGTTAGGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 933  
Db 1525 AATTGTTTCTAGTTAGGCGGCGGCTGTGGCTCGGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1584  
QY 934 -CAGCCATGGGCGGCTGTAGCAACTGTGAGCGGAGCGGCGGCTGAGCGGCGGCTC 991  
Db 1585 -CAGCCATGGGCGGCTGTAGCAACTGTGAGCGGAGCGGCGGCTGAGCGGCGGCTC 1644  
QY 992 CCA 994  
Db 1645 CCA 1647

RESULT 13  
AC008752/c  
LOCUS  
DEFINITION  
Homo sapiens chromosome 19 clone CTD-2623N2, complete sequence.  
ACCESSION  
AC008752  
138036 bp DNA linear PRI 13-JUL-2002



```

VERSION      AC008752.6  GI:10944453
KEYWORDS     Homo sapiens (human)
SOURCE       Homo sapiens
ORGANISM     Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 138036)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 138036)
AUTHORS      DOE Joint Genome Institute.
TITLE        Direct Submission
JOURNAL      Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE    3 (bases 1 to 138036)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE        Direct Submission
JOURNAL      Submitted (13-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
COMMENT      On Oct 21, 2000 this sequence version replaced gi:7711334.
              Draft sequence produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Finishing Completed at Stanford Human Genome Center
              www.shgc.stanford.edu
              Quality: Phrap Quality >=40 99.9% of Sequence;
              Estimated Total Number of Errors is 1.
              STS Content:
              SHGC-19162 G31077.

FEATURES             Location/Qualifiers
     source           1..138036
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTD-2623N2"

ORIGIN
Query Match      57.1%; Score 567.4; DB 9; Length 138036;
Best Local Similarity 98.5%; Pred. No. 9.9e-110;
Matches 594; Conservative 0; Mismatches 6; Indels 3; Gaps 2;

QY 395 CCTTCAGCAGAGTCAGATGCAGAGCCATTGTAAGACGCCCTCGTTTGGCTGCGGACGG 454
DB 5987 CTGGCTCCAGGTCCAGATGCAGAGCCATTGTAAGACGCCCTCGTTTGGCTGCGGACGG 5928

QY 455 GGGAGATGAGCGGCGCGGTGTTACGGATTCGGGATCCACATCATCTCCCGCACTGAT 514
DB 5927 GGGAGATGAGCGGCGCGGTGTTACGGATTCGGGATCCACATCATCTCCCGCACTGAT 5868

QY 515 GAGGGTGGGAGCCCGAGGCTCGGCTCGGGGAGGCGAGGCGGCTAGAGCCGCGAGGTC 574
DB 5867 GAGGGTGGGAGCCCGAGGCTCGGCTCGGGGAGGCGAGGCGGCTAGAGCCGCGAGGTC 5808

QY 575 CCCTTTGCCCGCCAGCAGTCGGCGGAAACCCCGCCACTCCCTTGCACCGTCACACAGTATT 634
DB 5807 CCCTTTGCCCGCCAGCAGTCGGCGGAAACCCCGCCACTCCCTTGCACCGTCACACAGTATT 5748

QY 635 ATTGTTCCCAATAGGCTGGAGGGGCGCTTCCAGATTGGGGGCGCTGGGGTCCCACT 694
DB 5747 ATTGTTCCCAATAGGCTGGAGGGGCGCTTCCAGATTGGGGGCGCTGGGGTCCCACT 5688

QY 695 CCTGTGTCATCCCGAGTGGGCTGGAGCCGAGATTCCTCCCTTAAGGAATTGACTTCA 754
DB 5687 CCTGTGTCATCCCGAGTGGGCTGGAGCCGAGATTCCTCCCTTAAGGAATTGACTTCA 5628

QY 755 GCAGGGGTGGGAGGCTCCAGAGCCAGGCGAGTGTGGTGGGAGGGGTGTTCCTCAAGAGAA 814
              |||||

```

```

DB 5627 GCAGGGGTGGGAGGCTCCAGAGCCAGAGCCAGGTGTGGTGGGAGGGGTGTTCCTCAAGAGAA 5568
QY 815 GGCTGTGTAGCAGAGCGCGCGCGTGTCCCGCCAGGTGTCTGGAGGCGAGACTCCGAGGGCCG 874
DB 5567 GGCTGTGTAGCAGAGCGCGCGCGTGTCCCGCCAGGTGTCTGGAGGCGAGACTCCGAGGGCCG 5508
QY 875 AATTGTTTCTAGTTAGGCCACGCTCTCTCTTCAGTGTGCAAAAGTGAACACTCATGCGG- 933
DB 5507 AATTGTTTCTAGTTAGGCCACGCTCTCTCTTCAGTGTGCAAAAGTGAACACTCATGCGG 5448
QY 934 -CAGCCATGGCGCTCTGAGCAACTGTGAG-ACCCTTTACCCCCCAATTAAACCCAGAA 991
DB 5447 CCAGCCATGGCGCTCTGAGCAACTGTGAGCAACTGTGAGCAACTGTGAGCAACTGTGAGCA 5388
QY 992 CCA 994
DB 5387 CCA 5385

RESULT 14
AC008742      194624 bp      DNA      linear      PRI 12-JUL-2002
LOCUS        Homo sapiens chromosome 19 clone CTD-2553C6, complete sequence.
DEFINITION   AC008742
ACCESSION    AC008742
VERSION      AC008742.9  GI:21734928
KEYWORDS     HTG.
SOURCE       Homo sapiens (human)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 194624)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 194624)
AUTHORS      DOE Joint Genome Institute.
TITLE        Direct Submission
JOURNAL      Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE    3 (bases 1 to 194624)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE        Direct Submission
JOURNAL      Submitted (25-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
REFERENCE    4 (bases 1 to 194624)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE        Direct Submission
JOURNAL      Submitted (12-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
COMMENT      On Jul 12, 2002 this sequence version replaced gi:21206217.
              Draft Sequence Produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Finishing Completed at Stanford Human Genome Center
              www.shgc.stanford.edu
              Quality: Phrap Quality >=40 99.9% of Sequence;
              Estimated Total Number of Errors is 0.1.
              Location/Qualifiers
              1..194624
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /chromosome="19"
              /clone="CTD-2553C6"

ORIGIN
Query Match      57.1%; Score 567.4; DB 9; Length 194624;
Best Local Similarity 98.5%; Pred. No. 9.6e-110;
Matches 594; Conservative 0; Mismatches 6; Indels 3; Gaps 2;

QY 395 CCTTCAGCAGAGTCAGATGCAGAGCCATTGTAAGACGCCCTCGTTTGGCTGCGGACGG 454
DB 16518 CTTGGCTCCAGGTCCAGATGCAGAGCCATTGTAAGACGCCCTCGTTTGGCTGCGGACGG 16577

```

QY	455	GGGAGATGAGCGGGCCGCTGTTTCACGATTCGCGGATCCACATCATCTCTCCGACTGAGT	514
Db	16578	GGGAGATGAGCGGGCCGCTGTTTCACGATTCGCGGATCCACATCATCTCTCCGACTGAGT	16637
QY	515	GAGGGTGGGAGCCAGAGCCCTGGGCTCGGGGCGAGGGCGGCGGTAGGCGGCGGAGCTC	574
Db	16638	GAGGGTGGGAGCCAGAGCCCTGGGCTCGGGGCGAGGGCGGCGGTAGGCGGCGGAGCTC	16697
QY	575	CCCTTGGCCCGCAGCCAGTGGCCGAAACCCCGCACTCCCTGCCACCGTCACACAGTATTT	634
Db	16698	CCCTTGGCCCGCAGCCAGTGGCCGAAACCCCGCACTCCCTGCCACCGTCACACAGTATTT	16757
QY	635	ATTGTTCCACAAATGCTGGGAGGGGGCCCTTCAGATTGGGGGCCCTGGGGTCCCGACT	694
Db	16758	ATTGTTCCACAAATGCTGGGAGGGGGCCCTTCAGATTGGGGGCCCTGGGGTCCCGACT	16817
QY	695	CCCTGTCCATCCCGAGTGGGGTGGCGGCGGAGGCTTCCCTTAAGGAATTTGACTTCA	754
Db	16818	CCCTGTCCATCCCGAGTGGGGTGGCGGCGGAGGCTTCCCTTAAGGAATTTGACTTCA	16877
QY	755	GCAAGGGTGGGAGGCTCCAGAGCCAGGCGAGTGTGGTGGGAGGGGTGTTCCAAAGAGAA	814
Db	16878	GCAAGGGTGGGAGGCTCCAGAGCCAGGCGAGTGTGGTGGGAGGGGTGTTCCAAAGAGAA	16937
QY	815	GGCTGTGTCAGCAGAGCGCGCCCGCTGTCCCGCCAGGTCCTGGAGGCGAGACTCGAGGGGCCG	874
Db	16938	GGCTGTGTCAGCAGAGCGCGCCCGCTGTCCCGCCAGGTCCTGGAGGCGAGACTCGAGGGGCCG	16997
QY	875	AATTGTTTCTAGTTAGGCCACGCTCTCTGTTTCAGTTCGCAAGGTGAACACTCATCGGC	933
Db	16998	AATTGTTTCTAGTTAGGCCACGCTCTCTGTTTCAGTTCGCAAGGTGAACACTCATCGGC	17057
QY	934	-CAGCCATGGGCGCTCTGAGCAACTGTGCAG-ACCCTTTTCACCCCAATTAACCCAGAA	991
Db	17058	CCAGCCATGGGCGCTCTGAGCAACTGTGCAGCAACCCTTTCACCCCAATTAACCCAGAA	17117
QY	992	CCA 994	
Db	17118	CCA 17120	
RESULT 15			
AB009691			
LOCUS	Mus sp. mRNA for PIN1, complete cds.		
DEFINITION	Mus sp. mRNA for PIN1, complete cds.		
ACCESSION	AB009691		
VERSION	AB009691.1 GI:6468199		
KEYWORDS	mpin1; PIN1.		
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Fujimori, F., Takahashi, K., Uchida, C. and Uchida, T.		
JOURNAL	Mice lacking Pin1 develop normally, but are defective in entering		
MEDLINE	cell cycle from G(0) arrest		
PUBMED	Biochem. Biophys. Res. Commun. 265 (3), 658-663 (1999)		
REFERENCE	20070807		
AUTHORS	10600477		
TITLE	2 (bases 1 to 954)		
JOURNAL	Fujimori, F. and Uchida, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (12-DEC-1997) Fumihiko Fujimori, Science University of		
TITLE	Tokyo, Department of Biological Science & Technology, Laboratory of		
JOURNAL	Genome Biology; Yamazaki, Noda, Chiba 278, Japan		
REFERENCE	(E-mail: fujimori@rs.noda.sut.ac.jp, Tel: +81-471-24-1501,		
AUTHORS	Fax: +81-471-25-1841)		
TITLE	Location/Qualifiers		
JOURNAL	1..954		
REFERENCE	/organism="Mus sp."		
AUTHORS	/mol_type="mRNA"		
TITLE	/db_xref="taxon:10095"		
JOURNAL	1..954		
FEATURES	source		
LOCUS	954 bp mRNA linear ROD 26-NOV-1999		
DEFINITION	Mus sp. mRNA for PIN1, complete cds.		
ACCESSION	AB009691		
VERSION	AB009691.1 GI:6468199		
KEYWORDS	mpin1; PIN1.		
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Fujimori, F., Takahashi, K., Uchida, C. and Uchida, T.		
JOURNAL	Mice lacking Pin1 develop normally, but are defective in entering		
MEDLINE	cell cycle from G(0) arrest		
PUBMED	Biochem. Biophys. Res. Commun. 265 (3), 658-663 (1999)		
REFERENCE	20070807		
AUTHORS	10600477		
TITLE	2 (bases 1 to 954)		
JOURNAL	Fujimori, F. and Uchida, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (12-DEC-1997) Fumihiko Fujimori, Science University of		
TITLE	Tokyo, Department of Biological Science & Technology, Laboratory of		
JOURNAL	Genome Biology; Yamazaki, Noda, Chiba 278, Japan		
REFERENCE	(E-mail: fujimori@rs.noda.sut.ac.jp, Tel: +81-471-24-1501,		
AUTHORS	Fax: +81-471-25-1841)		
TITLE	Location/Qualifiers		
JOURNAL	1..954		
REFERENCE	/organism="Mus sp."		
AUTHORS	/mol_type="mRNA"		
TITLE	/db_xref="taxon:10095"		
JOURNAL	1..954		
FEATURES	source		
LOCUS	954 bp mRNA linear ROD 26-NOV-1999		
DEFINITION	Mus sp. mRNA for PIN1, complete cds.		
ACCESSION	AB009691		
VERSION	AB009691.1 GI:6468199		
KEYWORDS	mpin1; PIN1.		
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Fujimori, F., Takahashi, K., Uchida, C. and Uchida, T.		
JOURNAL	Mice lacking Pin1 develop normally, but are defective in entering		
MEDLINE	cell cycle from G(0) arrest		
PUBMED	Biochem. Biophys. Res. Commun. 265 (3), 658-663 (1999)		
REFERENCE	20070807		
AUTHORS	10600477		
TITLE	2 (bases 1 to 954)		
JOURNAL	Fujimori, F. and Uchida, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (12-DEC-1997) Fumihiko Fujimori, Science University of		
TITLE	Tokyo, Department of Biological Science & Technology, Laboratory of		
JOURNAL	Genome Biology; Yamazaki, Noda, Chiba 278, Japan		
REFERENCE	(E-mail: fujimori@rs.noda.sut.ac.jp, Tel: +81-471-24-1501,		
AUTHORS	Fax: +81-471-25-1841)		
TITLE	Location/Qualifiers		
JOURNAL	1..954		
REFERENCE	/organism="Mus sp."		
AUTHORS	/mol_type="mRNA"		
TITLE	/db_xref="taxon:10095"		
JOURNAL	1..954		
FEATURES	source		
LOCUS	954 bp mRNA linear ROD 26-NOV-1999		
DEFINITION	Mus sp. mRNA for PIN1, complete cds.		
ACCESSION	AB009691		
VERSION	AB009691.1 GI:6468199		
KEYWORDS	mpin1; PIN1.		
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Fujimori, F., Takahashi, K., Uchida, C. and Uchida, T.		
JOURNAL	Mice lacking Pin1 develop normally, but are defective in entering		
MEDLINE	cell cycle from G(0) arrest		
PUBMED	Biochem. Biophys. Res. Commun. 265 (3), 658-663 (1999)		
REFERENCE	20070807		
AUTHORS	10600477		
TITLE	2 (bases 1 to 954)		
JOURNAL	Fujimori, F. and Uchida, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (12-DEC-1997) Fumihiko Fujimori, Science University of		
TITLE	Tokyo, Department of Biological Science & Technology, Laboratory of		
JOURNAL	Genome Biology; Yamazaki, Noda, Chiba 278, Japan		
REFERENCE	(E-mail: fujimori@rs.noda.sut.ac.jp, Tel: +81-471-24-1501,		
AUTHORS	Fax: +81-471-25-1841)		
TITLE	Location/Qualifiers		
JOURNAL	1..954		
REFERENCE	/organism="Mus sp."		
AUTHORS	/mol_type="mRNA"		
TITLE	/db_xref="taxon:10095"		
JOURNAL	1..954		
FEATURES	source		
LOCUS	954 bp mRNA linear ROD 26-NOV-1999		
DEFINITION	Mus sp. mRNA for PIN1, complete cds.		
ACCESSION	AB009691		
VERSION	AB009691.1 GI:6468199		
KEYWORDS	mpin1; PIN1.		
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Fujimori, F., Takahashi, K., Uchida, C. and Uchida, T.		
JOURNAL	Mice lacking Pin1 develop normally, but are defective in entering		
MEDLINE	cell cycle from G(0) arrest		
PUBMED	Biochem. Biophys. Res. Commun. 265 (3), 658-663 (1999)		
REFERENCE	20070807		
AUTHORS	10600477		
TITLE	2 (bases 1 to 954)		
JOURNAL	Fujimori, F. and Uchida, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (12-DEC-1997) Fumihiko Fujimori, Science University of		
TITLE	Tokyo, Department of Biological Science & Technology, Laboratory of		
JOURNAL	Genome Biology; Yamazaki, Noda, Chiba 278, Japan		
REFERENCE	(E-mail: fujimori@rs.noda.sut.ac.jp, Tel: +81-471-24-1501,		
AUTHORS	Fax: +81-471-25-1841)		
TITLE	Location/Qualifiers		
JOURNAL	1..954		
REFERENCE	/organism="Mus sp."		
AUTHORS	/mol_type="mRNA"		
TITLE	/db_xref="taxon:10095"		
JOURNAL	1..954		
FEATURES	source		
LOCUS	954 bp mRNA linear ROD 26-NOV-1999		
DEFINITION	Mus sp. mRNA for PIN1, complete cds.		
ACCESSION	AB009691		
VERSION	AB009691.1 GI:6468199		
KEYWORDS	mpin1; PIN1.		
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Fujimori, F., Takahashi, K., Uchida, C. and Uchida, T.		
JOURNAL	Mice lacking Pin1 develop normally, but are defective in entering		
MEDLINE	cell cycle from G(0) arrest		
PUBMED	Biochem. Biophys. Res. Commun. 265 (3), 658-663 (1999)		
REFERENCE	20070807		
AUTHORS	10600477		
TITLE	2 (bases 1 to 954)		
JOURNAL	Fujimori, F. and Uchida, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (12-DEC-1997) Fumihiko Fujimori, Science University of		
TITLE	Tokyo, Department of Biological Science & Technology, Laboratory of		
JOURNAL	Genome Biology; Yamazaki, Noda, Chiba 278, Japan		
REFERENCE	(E-mail: fujimori@rs.noda.sut.ac.jp, Tel: +81-471-24-1501,		
AUTHORS	Fax: +81-471-25-1841)		
TITLE	Location/Qualifiers		
JOURNAL	1..954		
REFERENCE	/organism="Mus sp."		
AUTHORS	/mol_type="mRNA"		
TITLE	/db_xref="taxon:10095"		
JOURNAL	1..954		
FEATURES	source		
LOCUS	954 bp mRNA linear ROD 26-NOV-1999		
DEFINITION	Mus sp. mRNA for PIN1, complete cds.		
ACCESSION	AB009691		
VERSION	AB009691.1 GI:6468199		
KEYWORDS	mpin1; PIN1.		
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Fujimori, F., Takahashi, K., Uchida, C. and Uchida, T.		
JOURNAL	Mice lacking Pin1 develop normally, but are defective in entering		
MEDLINE	cell cycle from G(0) arrest		
PUBMED	Biochem. Biophys. Res. Commun. 265 (3), 658-663 (1999)		
REFERENCE	20070807		
AUTHORS	10600477		
TITLE	2 (bases 1 to 954)		
JOURNAL	Fujimori, F. and Uchida, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (12-DEC-1997) Fumihiko Fujimori, Science University of		
TITLE	Tokyo, Department of Biological Science & Technology, Laboratory of		
JOURNAL	Genome Biology; Yamazaki, Noda, Chiba 278, Japan		
REFERENCE	(E-mail: fujimori@rs.noda.sut.ac.jp, Tel: +81-471-24-1501,		
AUTHORS	Fax: +81-471-25-1841)		
TITLE	Location/Qualifiers		
JOURNAL	1..954		
REFERENCE	/organism="Mus sp."		
AUTHORS	/mol_type="mRNA"		
TITLE	/db_xref="taxon:10095"		
JOURNAL	1..954		
FEATURES	source		
LOCUS	954 bp mRNA linear ROD 26-NOV-1999		
DEFINITION	Mus sp. mRNA for PIN1, complete cds.		
ACCESSION	AB009691		
VERSION	AB009691.1 GI:6468199		
KEYWORDS	mpin1; PIN1.		
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Fujimori, F., Takahashi, K., Uchida, C. and Uchida, T.		
JOURNAL	Mice lacking Pin1 develop normally, but are defective in entering		
MEDLINE	cell cycle from G(0) arrest		
PUBMED	Biochem. Biophys. Res. Commun. 265 (3), 658-663 (1999)		
REFERENCE	20070807		
AUTHORS	10600477		
TITLE	2 (bases 1 to 954)		
JOURNAL	Fujimori, F. and Uchida, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (12-DEC-1997) Fumihiko Fujimori, Science University of		
TITLE	Tokyo, Department of Biological Science & Technology, Laboratory of		
JOURNAL	Genome Biology; Yamazaki, Noda, Chiba 278, Japan		
REFERENCE	(E-mail: fujimori@rs.noda.sut.ac.jp, Tel: +81-471-24-1501,		
AUTHORS	Fax: +81-471-25-1841)		
TITLE	Location/Qualifiers		
JOURNAL	1..954		
REFERENCE	/organism="Mus sp."		
AUTHORS	/mol_type="mRNA"		
TITLE	/db_xref="taxon:10095"		
JOURNAL	1..954		
FEATURES	source		
LOCUS	954 bp mRNA linear ROD 26-NOV-1999		
DEFINITION	Mus sp. mRNA for PIN1, complete cds.		
ACCESSION	AB009691		
VERSION	AB009691.1 GI:6468199		
KEYWORDS	mpin1; PIN1.		
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Fujimori, F., Takahashi, K., Uchida, C. and Uchida, T.		
JOURNAL	Mice lacking Pin1 develop normally, but are defective in entering		
MEDLINE	cell cycle from G(0) arrest		
PUBMED	Biochem. Biophys. Res. Commun. 265 (3), 658-663 (1999)		
REFERENCE	20070807		
AUTHORS	10600477		
TITLE	2 (bases 1 to 954)		
JOURNAL	Fujimori, F. and Uchida, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (12-DEC-1997) Fumihiko Fujimori, Science University of		
TITLE	Tokyo, Department of Biological Science & Technology, Laboratory of		
JOURNAL	Genome Biology; Yamazaki, Noda, Chiba 278, Japan		
REFERENCE	(E-mail: fujimori@rs.noda.sut.ac.jp, Tel: +81-471-24-1501,		
AUTHORS	Fax: +81-471-25-1841)		
TITLE	Location/Qualifiers		
JOURNAL	1..954		
REFERENCE	/organism="Mus sp."		
AUTHORS	/mol_type="mRNA"		
TITLE	/db_xref="taxon:10095"		
JOURNAL	1..954		
FEATURES	source		
LOCUS	954 bp mRNA linear ROD 26-NOV-1999		
DEFINITION	Mus sp. mRNA for PIN1, complete cds.		
ACCESSION	AB009691		
VERSION	AB009691.1 GI:6468199		
KEYWORDS	mpin1; PIN1.		
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Fujimori, F., Takahashi, K., Uchida, C. and Uchida, T.		
JOURNAL	Mice lacking Pin1 develop normally, but are defective in entering		
MEDLINE	cell cycle from G(0) arrest		
PUBMED	Biochem. Biophys. Res. Commun. 265 (3), 658-663 (1999)		
REFERENCE	20070807		
AUTHORS	10600477		
TITLE	2 (bases 1 to 954)		
JOURNAL	Fujimori, F. and Uchida, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (12-DEC-1997) Fumihiko Fujimori, Science University of		
TITLE	Tokyo, Department of Biological Science & Technology, Laboratory of		
JOURNAL	Genome Biology; Yamazaki, Noda, Chiba 278, Japan		
REFERENCE	(E-mail: fujimori@rs.noda.sut.ac.jp, Tel: +81-471-24-1501,		
AUTHORS	Fax: +81-471-25-1841)		
TITLE	Location/Qualifiers		
JOURNAL	1..954		
REFERENCE	/organism="Mus sp."		
AUTHORS	/mol_type="mRNA"		
TITLE	/db_xref="taxon:10095"		
JOURNAL	1..954		
FEATURES	source		
LOCUS	954 bp mRNA linear ROD 26-NOV-1999		
DEFINITION	Mus sp. mRNA for PIN1, complete cds.		
ACCESSION	AB009691		
VERSION	AB009691.1 GI:6468199		
KEYWORDS	mpin1; PIN1.		
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Fujimori, F., Takahashi, K., Uchida, C. and Uchida, T.		
JOURNAL	Mice lacking Pin1 develop normally, but are defective in entering		
MEDLINE	cell cycle from G(0) arrest		
PUBMED	Biochem. Biophys. Res. Commun. 265 (3), 658-663 (1999)		
REFERENCE	20070807		
AUTHORS	10600477		
TITLE	2 (bases 1 to 954)		
JOURNAL	Fujimori, F. and Uchida, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (12-DEC-1997) Fumihiko Fujimori, Science University of		
TITLE	Tokyo, Department of Biological Science & Technology, Laboratory of		
JOURNAL	Genome Biology; Yamazaki, Noda, Chiba 278, Japan		
REFERENCE	(E-mail: fujimori@rs.noda.sut.ac.jp, Tel: +81-471-24-1501,		
AUTHORS	Fax: +81-471-25-1841)		
TITLE	Location/Qualifiers		
JOURNAL	1..954		
REFERENCE	/organism="Mus sp."		
AUTHORS	/mol_type="mRNA"		
TITLE	/db_xref="taxon:10095"		
JOURNAL	1..954		
FEATURES	source		
LOCUS	954 bp mRNA linear ROD 26-NOV-1999		
DEFINITION	Mus sp. mRNA for PIN1, complete cds.		
ACCESSION	AB009691		
VERSION	AB009691.1 GI:6468199		
KEYWORDS	mpin1; PIN1.		
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Fujimori, F., Takahashi, K., Uchida, C. and Uchida, T.		
JOURNAL	Mice lacking Pin1 develop normally, but are defective in entering		
MEDLINE	cell cycle from G(0) arrest		
PUBMED	Biochem. Biophys. Res. Commun. 265 (3), 658-663 (1999)		
REFERENCE	20070807		
AUTHORS	10600477		
TITLE	2 (bases 1 to 954)		
JOURNAL	Fujimori, F. and Uchida, T.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (12-DEC-1997) Fumihiko Fujimori, Science University of		
TITLE	Tokyo, Department of Biological Science & Technology, Laboratory of		
JOURNAL	Genome Biology; Yamazaki, Noda, Chiba 278, Japan		
REFERENCE	(E-mail: fujimori@rs.noda.sut.ac.jp, Tel: +81-471-24-1501,		
AUTHORS	Fax: +81-471-25-1841)		
TITLE	Location/Qualifiers		
JOURNAL	1..954		
REFERENCE	/organism="Mus sp."		
AUTHORS			



Tue Jun 29 08:37:21 2004

u49070.rge

Page 13

Search completed: June 28, 2004, 16:23:11  
Job time : 4369.63 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:24 ; Search time 545.075 Seconds  
(without alignments)  
7747.025 Million cell updates/sec

Title: U49070

Perfect score: 994

Sequence: 1 TCCTGGCCAGACCTCGAGG.....CCCAATTAAACCCAGAACCA 994

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Match	Length	DB ID	Description
1	994	100.0	994	5	Aah76459	Aah76459 cDNA corr
2	994	100.0	1014	2	Aat68888	Aat68888 NIVA-inte
3	892.6	89.8	3125	9	Adb63201	Adb63201 Human CDN
4	892.6	89.8	3143	4	Aad05229	Aad05229 Human sec
5	887.8	89.3	1143	6	Abk35196	Abk35196 Human CDN
6	729	73.3	1034	5	Aas76759	Aas76759 DNA encod
7	593.8	59.7	1045	5	Aas76134	Aas76134 DNA encod
8	492.2	49.5	580	2	Aaz33510	Aaz33510 Human pro
9	392	39.4	458	8	Ach32387	Ach32387 Human end
10	217.6	21.9	559	4	Ab62612	Ab62612 Human foe
11	217.6	21.9	559	4	Aai42606	Aai42606 Probe #11
12	217.6	21.9	559	4	Aak36818	Aak36818 Human bon
13	217.6	21.9	559	4	Aak10973	Aak10973 Human bra
14	217.6	21.9	559	6	AbS10823	AbS10823 Human gen
15	213	21.4	213	4	AbA75131	AbA75131 Human foe
16	213	21.4	213	4	Aai5672	Aai5672 Probe #24
17	213	21.4	213	4	Aak49765	Aak49765 Human bon
18	213	21.4	213	4	Aak23649	Aak23649 Human bra
19	213	21.4	213	6	AbS23259	AbS23259 Human gen
20	208.4	21.0	439	6	AbI99260	AbI99260 Mouse isc
21	203.2	20.4	314	4	Aaf92335	Aaf92335 Bovine ma
22	200.8	20.2	296	2	AAQ60623	AAQ60623 Human bra
23	178	17.9	601	9	Adb55060	Adb55060 Toxicity-

C	24	178	17.9	601	9	ADB49564	ADB49564 Primary r
	25	161.4	16.2	995	4	ABL18667	ABL18667 Drosophil
	26	112	11.3	468	5	AAS76133	AAS76133 DNA encod
	27	110.4	11.1	537	7	ABT19273	ABT19273 Aspergill
	28	110	11.1	345	5	AAS76758	AAS76758 DNA encod
	29	108.2	10.9	406	3	Aaf11336	Aaf11336 Aspergill
	30	107.2	10.8	537	7	ABT21093	ABT21093 Aspergill
C	31	98.4	9.9	3475	4	ABL18666	ABL18666 Drosophil
	32	98.4	9.9	5330	4	ABL03148	ABL03148 Drosophil
C	33	98.4	9.9	7208	4	ABL02698	ABL02698 Drosophil
	34	84.6	8.5	366	7	ABZ53193	ABZ53193 Aspergill
	35	84.4	8.5	285	6	ABN79537	ABN79537 Human iso
	36	80	8.0	833	3	AAC47563	AAC47563 Arabidops
	37	78.4	7.9	747	3	AAC33551	AAC33551 Arabidops
	38	78.4	7.9	835	3	AAC36658	AAC36658 Arabidops
	39	74.8	7.5	662	3	AAR07816	AAR07816 Fusarium
	40	67.6	6.8	630	9	ADB69628	ADB69628 C. neofo
	41	66	6.6	652	7	ABT18679	ABT18679 Aspergill
	42	66	6.6	652	7	ABT20495	ABT20495 Aspergill
	43	66	6.6	2652	7	ABT19899	ABT19899 Aspergill
	44	66	6.6	2652	7	ABT18085	ABT18085 Aspergill
C	45	66	6.6	5360	4	ABL20104	ABL20104 Drosophil

## ALIGNMENTS

RESULT 1  
AAH76459  
ID AAH76459 standard; cDNA; 994 BP.  
XX  
AC AAH76459;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE cDNA corresponding to human IFN-alpha induced gene encoding PIN-1.

XX Human; interferon-alpha induced gene; type I interferon treatment;  
XX chronic viral hepatitis; relapsing remitting multiple sclerosis;  
XX neoplastic disease; IFN-alpha; interferon-alpha; PIN-1; ss.  
XX

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 25..516

FT /\*tag= a

FT /product= "PIN-1"

XX WO200159155-A2.

XX 16-AUG-2001.

XX 09-FEB-2001; 2001WO-GB0000578.

XX 11-FEB-2000; 2000GB-00003203.

XX 11-FEB-2000; 2000GB-00003204.

XX 11-FEB-2000; 2000GB-00003205.

XX 11-FEB-2000; 2000GB-00003206.

XX 11-FEB-2000; 2000GB-00003207.

XX 11-FEB-2000; 2000GB-00003208.

XX 11-FEB-2000; 2000GB-00003210.

XX 11-FEB-2000; 2000GB-00003212.

XX 11-FEB-2000; 2000GB-00003213.

XX 11-FEB-2000; 2000GB-00003215.

XX 11-FEB-2000; 2000GB-00003216.

XX 11-FEB-2000; 2000GB-00003219.

XX 11-FEB-2000; 2000GB-00003220.

XX 11-FEB-2000; 2000GB-00003221.

XX 11-FEB-2000; 2000GB-00003222.

XX 17-FEB-2000; 2000GB-00003768.

XX (PHAR-) PHARMA PACIFIC PTY LTD.

XX

PI Meritet J, Dron M, Tovey MG;  
 XX WPI; 2001-483570/52.  
 DR P-PSDB; AAG66532.  
 XX  
 PT Predicting responsiveness of a patient to treatment with a type I  
 PT interferon comprising determining the level of induced proteins after  
 PT treatment with a type I interferon.  
 XX  
 XX Example 2; Page 60-61; 133pp; English.  
 XX  
 CC The invention relates to a method for predicting responsiveness of a  
 CC patient to treatment with a type I interferon. The method comprises  
 CC determining the level of one or more proteins with a 846, 164, 126, 598,  
 CC 98, 177, 761, 361, 941, 657, 817, 429, 473, 399, 285 or 303 amino acid  
 CC sequence fully defined in the specification after treatment with a type I  
 CC interferon. The method allows a physician to determine whether a patient  
 CC suffering from chronic viral hepatitis, neoplastic disease or relapsing  
 CC remitting multiple sclerosis will respond favourably to Type I interferon  
 CC treatment via oromucosal administration. The present sequence is a cDNA  
 CC corresponding to an interferon-alpha induced gene that encodes one of the  
 CC polypeptides listed above  
 XX  
 XX Sequence 994 BP; 202 A; 309 C; 324 G; 159 T; 0 U; 0 Other;  
 SQ

Query Match 100.0%; Score 994; DB 5; Length 994;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-230;  
 Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGCCAGCAGCTCGAGGGAAGATGGCGGAGAGAGCTGCGCGCGGCTGGGAG 60  
 DB 1 TGTGTGCCAGCAGCTCGAGGGAAGATGGCGGAGAGAGCTGCGCGCGGCTGGGAG 60

QY 61 AAGCGCATGAGCGGAGCTCAGCGAGTGTACTTCAACACATCACTAACGCCAGC 120  
 DB 61 AAGCGCATGAGCGGAGCTCAGCGAGTGTACTTCAACACATCACTAACGCCAGC 120

QY 121 CAGTGGAGGGGCGGAGCGGAGCAGCAGTGTGGGCAAAACGGGAGGGAGCCT 180  
 DB 121 CAGTGGAGGGGCGGAGCGGAGCAGCAGTGTGGGCAAAACGGGAGGGAGCCT 180

QY 181 GCCAGGGTCCGCTGCTCGACCTGCTGGTGAAGCAGCAGCGGCGGCTCGTCC 240  
 DB 181 GCCAGGGTCCGCTGCTCGACCTGCTGGTGAAGCAGCAGCGGCGGCTCGTCC 240

QY 241 TGGCGGAGGAGAGATCACCCGGACAGAGGAGGCGCTGGAGCTGATCAACGGCTAC 300  
 DB 241 TGGCGGAGGAGAGATCACCCGGACAGAGGAGGCGCTGGAGCTGATCAACGGCTAC 300

QY 301 ATCCAGAGATCAAGTCGGGAGAGGAGCTTTGAGTCTCTGCGCTCACAGTTTCAGCGAC 360  
 DB 301 ATCCAGAGATCAAGTCGGGAGAGGAGCTTTGAGTCTCTGCGCTCACAGTTTCAGCGAC 360

QY 361 TGCAGCTCAGCAGGCGGAGAGCCTGGTGCCTTTCAGCAGAGTCAAGTGCAGAAG 420  
 DB 361 TGCAGCTCAGCAGGCGGAGAGCCTGGTGCCTTTCAGCAGAGTCAAGTGCAGAAG 420

QY 421 CCATTGAGAGCGCTGTTTGGCTCGGACCGGGAGATGAGCGGCGGCTGTTCAAG 480  
 DB 421 CCATTGAGAGCGCTGTTTGGCTCGGACCGGGAGATGAGCGGCGGCTGTTCAAG 480

QY 481 GATTCCGGGATCCACATCATCTCCGACTGAGTGGGAGGCGGAGGCGGCTGCGCT 540  
 DB 481 GATTCCGGGATCCACATCATCTCCGACTGAGTGGGAGGCGGAGGCGGCTGCGCT 540

QY 541 CGGGGAGGCGGAGGCGGCTAGGCGGCGGAGCTCCCGCTTCCCGCGCAGCCAGTGGCGGA 600  
 DB 541 CGGGGAGGCGGAGGCGGCTAGGCGGCGGAGCTCCCGCTTCCCGCGCAGCCAGTGGCGGA 600

QY 601 ACCCGGAGGCGGAGGCGGCTAGGCGGCGGAGCTTATTTGTTCCCAATGCTGGGAGGG 660  
 DB 601 ACCCGGAGGCGGAGGCGGCTAGGCGGCGGAGCTTATTTGTTCCCAATGCTGGGAGGG 660

QY 661 GCCCTTCAGATTGGGGGCGCTCCGACCTCCCTGTCCATCCCGAGTTGGGGTGC 720  
 DB 661 GCCCTTCAGATTGGGGGCGCTCCGACCTCCCTGTCCATCCCGAGTTGGGGTGC 720

QY 721 GACCGCCAGATTCTCCCTTTAAGGAATTGACTTCAGCAGGGGTGGAGGCTCCAGACCCA 780  
 DB 721 GACCGCCAGATTCTCCCTTTAAGGAATTGACTTCAGCAGGGGTGGAGGCTCCAGACCCA 780

QY 781 GGGCAGTGTGGAGGGGTGTTCCAAAGAGAAGGCTGTCTCAGCAGAGCGGCCCTG 840  
 DB 781 GGGCAGTGTGGAGGGGTGTTCCAAAGAGAAGGCTGTCTCAGCAGAGCGGCCCTG 840

QY 841 TCCCGCCAGTGTGGAGGCGAGCTTCGAGGCGCGAATTGTTTCTAGTTAGGCCAGCTCC 900  
 DB 841 TCCCGCCAGTGTGGAGGCGAGCTTCGAGGCGCGAATTGTTTCTAGTTAGGCCAGCTCC 900

QY 901 TCTGTTAGTTCGCAAAAGTGAACACTCATCGGCGAGCCTATGGGCGCTTGAGCAACTGTG 960  
 DB 901 TCTGTTAGTTCGCAAAAGTGAACACTCATCGGCGAGCCTATGGGCGCTTGAGCAACTGTG 960

QY 961 CAGACCTTTTCAACCCCAATTAAACCCAGAACCA 994  
 DB 961 CAGACCTTTTCAACCCCAATTAAACCCAGAACCA 994

RESULT 2  
 AAT68888  
 ID AAT68888 standard; DNA; 1014 BP.  
 XX  
 AC AAT68888;  
 XX  
 DT 09-DEC-1997 (first entry)  
 XX  
 DE NIMA-interacting protein Pin1 DNA.  
 XX  
 KW Pin1; protein interacting with NIMA; cell proliferation; mitosis;  
 KW peptidyl-propyl cis/trans isomerase; adenocarcinoma; cancer; leukaemia;  
 KW psoriasis; pemphigus vulgaris; rheumatoid arthritis;  
 KW acute respiratory distress syndrome; septic shock; inflammation; therapy;  
 XX ss.  
 XX Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 XX 25..516  
 XX CDS /\*tag= a  
 XX  
 XX W09717986-A1.  
 XX  
 XX 22-MAY-1997.  
 XX  
 XX 28-OCT-1996; 96WO-US017334.  
 XX  
 XX 13-NOV-1995; 95US-00555912.  
 XX  
 XX (SALK ) SALK INST BIOLOGICAL STUDIES.  
 XX  
 XX Hunter T, Lu KP;  
 XX  
 XX WPI; 1997-289057/26.  
 XX P-PSDB; AAW18312.  
 XX  
 XX Protein, Pin1, interacting with NIMA - used for treating cell  
 XX proliferative disorders.  
 XX  
 XX Claim 8; Page 49-50; 73pp; English.  
 XX  
 XX This DNA sequence codes for human Pin1 (AAW18312), an 18 kDa protein that  
 XX has peptidyl-propyl cis/trans isomerase activity, associates with NIMA  
 XX protein kinase, inhibits the mitosis-promoting function of NIMA when  
 XX overexpressed, and induces mitotic arrest and nuclear fragmentation when  
 XX depleted. It was identified using a yeast two-hybrid system with  
 XX Aspergillus nidulans nima as bait and cDNA from an HeLa library. A

CC recombinant expression vector comprising the DNA sequence and host cells  
 CC containing the vector are claimed. Methods are also claimed for  
 CC identifying proteins that inhibit the mitosis promoting function of NIMA  
 CC protein kinase and for controlling the growth of a cell by reducing Pini  
 CC activity or Pini expression using an inhibitor, anti-Pini antibody,  
 CC antisense nucleotide sequence or ribozyme, or by increasing Pini activity  
 CC in the presence of an activator or increasing Pini expression using an  
 CC enhancer. This allows treatment of cell proliferation disorders such  
 CC adenocarcinomas, cancers, psoriasis, pemphigus vulgaris, acute  
 CC respiratory distress syndrome, rheumatoid arthritis, septic shock and  
 CC inflammation  
 XX

SQ Sequence 1014 BP; 220 A; 310 C; 324 G; 160 T; 0 U; 0 Other;

Query Match 100.0%; Score 994; DB 2; Length 1014;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-230;  
 Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCTGGCCAGACCTCGAGGGAAGATGCGGACGAGGAGAGCTGCCGCCCGCTGGGAG 60  
 DB 1 TGCTGGCCAGACCTCGAGGGAAGATGCGGACGAGGAGAGCTGCCGCCCGCTGGGAG 60  
 QY 61 AAGCGCATGAGCCGAGCTCAGGCGAGTGTACTTCAACCAATCACTAAGCCAGC 120  
 DB 61 AAGCGCATGAGCCGAGCTCAGGCGAGTGTACTTCAACCAATCACTAAGCCAGC 120  
 QY 121 CAGTGGGAGCGCCAGCGGACACAGCAGCAGTGTGGCAAAAGGGGAGGGGCGCT 180  
 DB 121 CAGTGGGAGCGCCAGCGGACACAGCAGCAGTGTGGCAAAAGGGGAGGGGCGCT 180  
 QY 181 GCCAGGGTCCGCTGCTCGCACTCTGCTGTAAGCACAGCCAGTCAACGGCGGCTCGTCC 240  
 DB 181 GCCAGGGTCCGCTGCTCGCACTCTGCTGTAAGCACAGCCAGTCAACGGCGGCTCGTCC 240  
 QY 241 TGGCGGCGAGGAGAGATCACCCGGACCAAGGAGGAGGCCCTGGAGCTGATCAAGCGCTAC 300  
 DB 241 TGGCGGCGAGGAGAGATCACCCGGACCAAGGAGGAGGCCCTGGAGCTGATCAAGCGCTAC 300  
 QY 301 ATCCAGAGATCAAGTCGGGAGAGAGACATTTGAGTCTCTGGGCTCAGTTCAGCCAC 360  
 DB 301 ATCCAGAGATCAAGTCGGGAGAGAGACATTTGAGTCTCTGGGCTCAGTTCAGCCAC 360  
 QY 361 TGCAGCTCAGCAAGCCAGCGGAGACCTGGGTGCTTCAGCAGAGGTGAGTGCAGAG 420  
 DB 361 TGCAGCTCAGCAAGCCAGCGGAGACCTGGGTGCTTCAGCAGAGGTGAGTGCAGAG 420  
 QY 421 CCATTTGAAGCGCTCGTTTGGCTGCGGACGCGGAGAGATGAGCGGGCCGCTGTTACG 480  
 DB 421 CCATTTGAAGCGCTCGTTTGGCTGCGGACGCGGAGAGATGAGCGGGCCGCTGTTACG 480  
 QY 481 GATTCGGGATCCACATCATCTCCGCACTGAGTGTGAGGTGGGAGCCAGGCGCTGGCT 540  
 DB 481 GATTCGGGATCCACATCATCTCCGCACTGAGTGTGAGGTGGGAGCCAGGCGCTGGCT 540  
 QY 541 CGGGCAGGCGAGCGGCTAGCGCGGCGAGCTCCCTTGCCTGCCAGCCAGTGCAGC 600  
 DB 541 CGGGCAGGCGAGCGGCTAGCGCGGCGAGCTCCCTTGCCTGCCAGCCAGTGCAGC 600  
 QY 601 ACCCCCCACTCCCTGCCACCGTACACAGTATTTATTTGTTCCCAATGGCTGGAGGGG 660  
 DB 601 ACCCCCCACTCCCTGCCACCGTACACAGTATTTATTTGTTCCCAATGGCTGGAGGGG 660  
 QY 661 GCCCTTCCAGATTGGGGGCGCTGGGCTCCCACTCTCTGCTCCATCCCAAGTTGGGGCTGC 720  
 DB 661 GCCCTTCCAGATTGGGGGCGCTGGGCTCCCACTCTCTGCTCCATCCCAAGTTGGGGCTGC 720  
 QY 721 GACCGCCAGATTCTCCCTTAAAGAAATGACTTCAGCAGGGGTGGAGGCTCCAGACCCA 780  
 DB 721 GACCGCCAGATTCTCCCTTAAAGAAATGACTTCAGCAGGGGTGGAGGCTCCAGACCCA 780  
 QY 781 GGGCAGTGTGGAGGGGCTGTTCAAGAGAGAGGCTGGTCAGCAGAGCGCCCGCTG 840  
 DB 781 GGGCAGTGTGGAGGGGCTGTTCAAGAGAGAGGCTGGTCAGCAGAGCGCCCGCTG 840

QY 841 TCCCCCAGGTGCTGGAGGAGAGTCTGAGGCGCGAATTGTTCTAGTTAGGCCACGCTCC 900  
 DB 841 TCCCCCAGGTGCTGGAGGAGAGTCTGAGGCGCGAATTGTTCTAGTTAGGCCACGCTCC 900  
 QY 901 TCTGTTTCTGAGTGCAGGAGGTGAACACTCATGCGGAGCATGGGCGCTCTGAGCAACTGTG 960  
 DB 901 TCTGTTTCTGAGTGCAGGAGGTGAACACTCATGCGGAGCATGGGCGCTCTGAGCAACTGTG 960  
 QY 961 CAGACCTTTTACCCCCCAATTAACCCAGAACCA 994  
 DB 961 CAGACCTTTTACCCCCCAATTAACCCAGAACCA 994

## RESULT 3

ADB63201

ID ADB63201 standard; cDNA; 3125 BP.

XX ADB63201;

XX AC ADB63201;

XX DT 04-DEC-2003 (first entry)

XX DE Human cDNA encoding clone SPLEN20136700.

XX KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
 KW tissue regeneration; cell regeneration; membrane protein;  
 KW signal transduction-related protein; transcription-related protein;  
 KW osteoporosis; neurological disease; cancer; tumour.

XX OS Homo sapiens.

XX XX Key Location/Qualifiers

XX CDS 1296..1628

XX FT /\*tag= a

XX FT /product= "Clone SPLEN20136700 protein"

XX PN EPI308459-A2.

XX PD 07-MAY-2003.

XX PF 28-MAR-2002; 2002EP-00007401.

XX PR 05-NOV-2001; 2001JP-00379298.

XX PR 25-JAN-2002; 2002US-00350978.

XX XX (HELI-) HELIX RES INST.

XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.

XX PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

XX PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

XX PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

XX DR WPI; 2003-450961/43.

XX DR P-PSDB; ADB65171.

XX New polynucleotides and polypeptides, useful for developing a diagnostic  
 PT marker or medicines for regulation of their expression and activity, or  
 PT as targets of gene therapy.

XX PS Claim 1; Page; 222pp; English.

XX The invention discloses a polynucleotide comprising a sequence selected  
 CC from 1970 fully defined nucleotide sequences which encode novel  
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 CC or its partial peptide, an antibody binding to the polypeptide or peptide  
 CC of the polynucleotide, immunologically assaying the polypeptide or  
 CC peptide of the polynucleotide by contacting the polypeptide or peptide  
 CC with the antibody of the encoded protein, and observing the binding  
 CC between the two, a transformant carrying the polynucleotide in an  
 CC expressible manner and an antisense polynucleotide. The oligonucleotide  
 CC is useful as a primer for synthesising the polynucleotide, or as a probe  
 CC for detecting the polynucleotide. The polynucleotides and encoded  
 CC proteins are useful as pharmaceutical agents and many disease-related

CC Genes may be included in them, for developing a diagnostic marker or  
CC medicines for regulation of their expression and activity, or as targets  
CC of gene therapy. The genes are involved in tissue and/or cell  
CC regeneration. Membrane proteins, signal transduction-related proteins,  
CC transcription-related proteins, disease-related proteins and genes  
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate  
CC the activity or expression of the encoded protein to treat diseases. The  
CC sequence presented is a cDNA of the invention. Note: Some of the sequence  
CC data for this patent is not represented in the printed specification, but  
CC is based on sequence information supplied by the European Patent Office.

XX	Sequence	3125 BP; 622 A; 893 C; 805 G; 805 T; 0 U; 0 Other;
XX	Query Match	89.8%; Score 892.6; DB 9; Length 3125;
XX	Best Local Similarity	98.7%; Pred. No. 2.9e-205;
XX	Matches	921; Conservative 0; Mismatches 9; Indels 3; Gaps 2;
QY	65	GCATGAGCCGACGCTCAGGCGGAGTGTACTTCAACACATCACTAAACGCCAGCCAGT 124
DB	2191	GCCTGCTCCCTCCCTCCAGGCGGAGTGTACTTCAACACATCACTAAACGCCAGCCAGT 2250
QY	125	GGAGGCGGCCAGCGCAACAGCAGCAGTGTGTGGCAAAACGGGCGAGGGGAGCCCTGCCA 184
DB	2251	GGAGGCGGCCAGCGCAACAGCAGCAGTGTGTGGCAAAACGGGCGAGGGGAGCCCTGCCA 2310
QY	185	GGTCCGCTGTCTGCACTCTCTGTGAAGCAGCAGTCAAGCGGCGCTCTCTCTGGC 244
DB	2311	GGTCCGCTGTCTGCACTCTCTGTGAAGCAGCAGTCAAGCGGCGCTCTCTCTGGC 2370
QY	245	GGCAGGAGAGATCAACCCGACCAAGAGGAGGCGCTTGGAGCTGATCAACGGGTACATCC 304
DB	2371	GGCAGGAGAGATCAACCCGACCAAGAGGAGGCGCTTGGAGCTGATCAACGGGTACATCC 2430
QY	305	AGAAGATCAAGTGGGAGAGAGACTTTTCAGTCTCTGGCTCAAGTTTCAGGCACTGCA 364
DB	2431	AGAAGATCAAGTGGGAGAGAGACTTTTCAGTCTCTGGCTCAAGTTTCAGGCACTGCA 2490
QY	365	GCTCAGCAAGGCGAGGGAGACTGGTGTCTTTCAGCAGTTCAGTTCAGAGGCGAT 424
DB	2491	GCTCAGCAAGGCGAGGGAGACTGGTGTCTTTCAGCAGTTCAGTTCAGAGGCGAT 2550
QY	425	TTGAGAGCGCTCTGTTTGGCTGGGAGGAGATGAGCGGCGCTGTTCACGGATT 484
DB	2551	TTGAGAGCGCTCTGTTTGGCTGGGAGGAGATGAGCGGCGCTGTTCACGGATT 2610
QY	485	CCGCGATCCATCATCTCTCGCACTGAGTGAGGCTGGGAGGCGCCAGGCTGGCTGGG 544
DB	2611	CCGCGATCCATCATCTCTCGCACTGAGTGAGGCTGGGAGGCGCCAGGCTGGCTGGG 2670
QY	545	GCAGGCGAGGCGGCTAGGCGGCGAGCTCCCTTTCGCGCAGGAGTGGCGGAGCC 604
DB	2671	GCAGGCGAGGCGGCTAGGCGGCGAGCTCCCTTTCGCGCAGGAGTGGCGGAGCC 2730
QY	605	CCCACTCCCTGCCACCTGCACAGTATTTATTTTCCCAATGGCTGGGAGGCGGCC 664
DB	2731	CCCACTCCCTGCCACCTGCACAGTATTTATTTTCCCAATGGCTGGGAGGCGGCC 2790
QY	665	TTCCAGATTGGGGCGCTGGGCTCCCACTCCCTTTCAGTTCATCCAGTTGGGCTGGAC 724
DB	2791	TTCCAGATTGGGGCGCTGGGCTCCCACTCCCTTTCAGTTCATCCAGTTGGGCTGGAC 2850
QY	725	GCCAGATTCTCCCTTAAGGAATTCAGTTCAGCAGGGTGGAGGCTCCAGACCCAGGGC 784
DB	2851	GCCAGATTCTCCCTTAAGGAATTCAGTTCAGCAGGGTGGAGGCTCCAGACCCAGGGC 2910
QY	785	AGTGTGGTGGAGGGTGTTCCTCAAGAGAGGCGCTGGTTCAGCAGAGCGCCCTGTGCC 844
DB	2911	AGTGTGGTGGAGGGTGTTCCTCAAGAGAGGCGCTGGTTCAGCAGAGCGCCCTGTGCC 2970
QY	845	CCAGGCTCTGGAGGAGCAGTTCAGGCGGCGAATTTCTAGTTAGGCGAGCTCTCTG 904
DB	2971	CCAGGCTCTGGAGGAGCAGTTCAGGCGGCGAATTTCTAGTTAGGCGAGCTCTCTG 3030

QY	905	TTCACTGCGAAAGGTGAACACTCATGGG--CAGCCATGGGCGCTCTGAGCAACTGTGCA 962
DB	3031	TTCACTGCGAAAGGTGAACACTCATGGGCGCCAGCCATGGGCGCTCTGAGCAACTGTGCA 3090
QY	963	G-ACCTTTTACCCCCCAATTAAACCCAGAACCA 994
DB	3091	GCACCTTTACCCCCCAATTAAACCCAGAACCA 3123
RESULT 4		
AAD05229		
ID	AAD05229	standard; cDNA; 3143 BP.
XX	AC	AAD05229;
XX	DT	18-JUL-2001 (first entry)
XX	DE	Human secreted protein-encoding gene 10 cDNA clone HHEPJ23, SEQ ID NO:20.
XX	KW	Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
XX	KW	foetal abnormality; developmental abnormality; haematopoietic disorder;
XX	KW	immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX	KW	Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
XX	KW	psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
XX	KW	inflammation; neurological disorder; Alzheimer's disease; food additive;
XX	KW	angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
XX	KW	pregnancy-related disorder; endocrine disorder; infection; wound healing;
XX	KW	cell culture; chemotaxis; vulvular; binding partner identification;
XX	OS	Homo sapiens.
XX	PH	Key
FT	CDS	Location/Qualifiers
FT		/tag= a
FT		/product= "Human secreted protein"
FT	sig_peptide	175..237
FT		/tag= b
FT	mat_peptide	238..360
FT		/tag= c
FT		/product= "Mature human secreted protein"
XX		WO200134629-A1.
XX	PN	17-MAY-2001.
XX	PD	
XX	PF	08-NOV-2000; 2000WO-US030654.
XX	PR	12-NOV-1999; 99US-0164835P.
XX	PR	27-JUL-2000; 2000US-0221142P.
XX	PA	(HUMA-) HUMAN GENOME SCI INC.
XX	PI	Ruben SM, Komatsoulis GA, Wei P, Fiscella M, Baker KP;
XX	DR	WPI; 2001-308779/32.
XX	DR	P-PSDB; AAE01361.
XX	PT	New nucleic acid encoding one of 21 human secreted proteins for
XX	PT	diagnosing, preventing, treating or ameliorating medical conditions, such
XX	PT	as autoimmune disease and cancer, and used as a food additive or
XX	PT	preservative.
XX	PS	Claim 1; Page 386-387; 490pp; English.
XX	CC	AAD05220-AAD05282 represent cDNAs corresponding to 21 human secreted
XX	CC	protein genes, and AAE01352-AAE01413 represent the proteins they encode.
XX	CC	AAE01415-AAE01433 represent human secreted protein fragments or variants.
XX	CC	The secreted proteins and their genes are useful for preventing, treating
XX	CC	or ameliorating medical conditions, e.g., by protein or gene therapy.
XX	CC	Pathological conditions can be diagnosed by determining the amount of the
XX	CC	new protein in a sample or by determining the presence of mutations in

the new genes. Specific uses are described for each of the 21 genes.  
based on the tissues in which they are most highly expressed, and include  
developing products for the diagnosis or treatment of proliferative  
disorders, cancer, tumours, foetal and developmental abnormalities,  
haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
psoriasis), sepsis, atherosclerosis, cardiovascular disorders,  
angiogenic disorders, kidney disorders, gastrointestinal disorders,  
pregnancy-related disorders, endocrine disorders, and infectious. The  
proteins can also be used to aid wound healing and epithelial cell  
proliferation, to prevent skin aging due to sunburn, to maintain organs  
before transplantation, for supporting cell culture of primary tissues,  
to regenerate tissues, to identify their cognate ligands or binding  
partners, and in chemotaxis, and can be used as a food additive or  
preservative to modify storage properties. Antibodies specific for a  
protein of the invention can be used in alleviating symptoms associated  
with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
present sequence represents a human secreted protein-encoding cDNA of the  
invention

Sequence 3143 BP; 678 A; 888 C; 802 G; 762 T; 0 U; 13 Other;  
Query Match 89.8%; Score 892.6; DB 4; Length 3143;  
Best Local Similarity 98.7%; Pred. No. 3e-205;  
Matches 921; Conservative 0; Mismatches 9; Indels 3; Gaps 2;

QY 65 GCATGAGCCGAGCTCAGGCGAGTGACTACTTCAACCAACATCACTAAAGCCAGCCAGT 124  
DB 2089 GCCTGCCCTCCCTCCAGGCGAGTGACTACTTCAACCAACATCACTAAAGCCAGCCAGT 2148  
QY 125 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGTGCAAAACGCGCAGGGGAGCTCGCA 184  
DB 2149 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGTGCAAAACGCGCAGGGGAGCTCGCA 2208  
QY 185 GGGTCCGCTGTCTGCACCTCTGTGTGAAGCAGCAGTGTGTGCGGCGCCCTGCTGGC 244  
DB 2209 GGGTCCGCTGTCTGCACCTCTGTGTGAAGCAGCAGTGTGTGCGGCGCCCTGCTGGC 2268  
QY 245 GGCAGGAGAGATCAACCGGACCAAGGAGGAGGCGCTCGAGCTGATCAACGGCTACATCC 304  
DB 2269 GGCAGGAGAGATCAACCGGACCAAGGAGGAGGCGCTCGAGCTGATCAACGGCTACATCC 2328  
QY 305 AGAAGATCAAGTCGGGAGAGAGGACTTTGAGTCTCTGGCCCTCAGAGTTCAGGAGTCA 364  
DB 2329 AGAAGATCAAGTCGGGAGAGAGGACTTTGAGTCTCTGGCCCTCAGAGTTCAGGAGTCA 2388  
QY 365 GCTCAGCCAGGCGGAGGAGCCTGGGTGCTTCAGCAGAGGTCAGATGAGAGCCAT 424  
DB 2389 GCTCAGCCAGGCGGAGGAGCCTGGGTGCTTCAGCAGAGGTCAGATGAGAGCCAT 2448  
QY 425 TTGAAGAGCCTCTGTTTGGCTGCGGAGGAGAGTGTGTGTTTCAAGGAT 484  
DB 2449 TTGAAGAGCCTCTGTTTGGCTGCGGAGGAGAGTGTGTGTTTCAAGGAT 2508  
QY 485 CCGGATCAATCATCTCTCCGACATGAGTGGGTGGGAGCCAGGCTGGCTCGG 544  
DB 2509 CCGGATCAATCATCTCTCCGACATGAGTGGGTGGGAGCCAGGCTGGCTCGG 2568  
QY 545 GCAGGCGGCGGCTAGGCGGCGAGCTCCCTTGGCCGCGAGCAGTGGCCGAAACC 604  
DB 2569 GCAGGCGGCGGCTAGGCGGCGAGCTCCCTTGGCCGCGAGCAGTGGCCGAAACC 2628  
QY 605 CCCACTCTCCCTGCGACCGTCACACAGTATTTATTTGTTCCACAAATGCTGGGAGGGGCC 664  
DB 2629 CCCACTCTCCCTGCGACCGTCACACAGTATTTATTTGTTCCACAAATGCTGGGAGGGGCC 2688  
QY 665 TTCAGATTTGGGGGCGCTGGGGTCCCACTCCCTGTCCATCCCACTGGGGTGGGAC 724  
DB 2689 TTCAGATTTGGGGGCGCTGGGGTCCCACTCCCTGTCCATCCCACTGGGGTGGGAC 2748

QY 725 GCCAGATTCTCCCTTAAGGAATTGACTTTCAGCAGGGGTGGAGGCTCCAGACCCAGGGC 784  
DB 2749 GCCAGATTCTCCCTTAAGGAATTGACTTTCAGCAGGGGTGGAGGCTCCAGACCCAGGGC 2808  
QY 785 AGTGTGGTGGAGGGGTGTTCAAAGAGAAGGCTGTGTGTCAGCAGAGCCGCCCGTGTCCC 844  
DB 2809 AGTGTGGTGGAGGGGTGTTCAAAGAGAAGGCTGTGTGTCAGCAGAGCCGCCCGTGTCCC 2868  
QY 845 CCCAGGTGCTGGAGGCGAGCTCGAGGCGCGAATTGTTTCTAGTTAGGCCACGCTCCTCTG 904  
DB 2869 CCCAGGTGCTGGAGGCGAGCTCGAGGCGCGAATTGTTTCTAGTTAGGCCACGCTCCTCTG 2928  
QY 905 TTCAGTGCAGAAAGGTGAACACTCATCGG--CAGCCATGGCGCCCTCTGAGCAATGTGCA 962  
DB 2929 TTCAGTGCAGAAAGGTGAACACTCATCGGCGCCAGCCATGGCGCCCTCTGAGCAATGTGCA 2988  
QY 963 G-ACCTTTTACCCCCCAATTAACCCAGAACCA 994  
DB 2989 GCACCCCTTTACCCCCCAATTAACCCAGAACCA 3021

RESULT 5  
ABK35196  
ID ABK35196 standard; cDNA; 1143 BP.  
XX  
AC ABK35196;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Human cDNA encoding secreted protein #334.  
XX  
KW Human; secreted protein; gene; ss; nutritional supplement; haemophilia;  
KW viral infection; bacterial infection; fungal infection; diabetes; asthma;  
KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;  
KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;  
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;  
KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;  
KW tissue regeneration; wound healing; burn; haematopoiesis;  
KW myeloid cell deficiency; lymphoid cell deficiency.  
OS Homo sapiens.  
XX  
FN WO200177288-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 29-MAR-2001; 2001WO-US010224.  
XX  
PR 06-APR-2000; 2000US-0195582P.  
XX  
(GEMY ) GENETICS INST INC.  
XX  
PI Wong GS, Clark HF, Fechtel K, Agoetino MJ, Howes SH, Resnick RJ;  
PI Gulukota K, Graham JR;  
XX  
WPI; 2002-179321/23.  
XX  
PT Five hundred and ninety two polynucleotides derived from a variety of  
PT human tissue sources which encode secreted proteins, useful for treating  
PT immune deficiencies and disorders such as autoimmune disorders.  
XX  
PS Claim 1; Page 246; 372pp; English.  
XX  
CC The invention relates to 592 polynucleotides which have been derived from  
CC a variety of human tissue sources and which encode novel secreted  
CC proteins. The polynucleotides can be used as probes for the  
CC identification and isolation of full length cDNA and genomic DNA. The  
CC polynucleotides and proteins can also be used as nutritional supplements.  
CC The proteins are useful in the treatment of various immune deficiencies  
CC and disorders such as viral infections, bacterial infections, fungal  
CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple  
CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions  
CC and conditions (e.g. asthma). They are also useful for treating

CC	neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's disease), liver fibrosis, coagulation disorders (e.g. haemophilia), CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also useful for tissue regeneration, for wound healing and in the treatment of CC burns, incisions and ulcers. The proteins are also useful for regulating CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies. CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention
XX	Sequence 1143 BP; 235 A; 359 C; 343 G; 206 T; 0 U; 0 Other;
SQ	Query Match 89.3%; Score 887.8; DB 6; Length 1143; Best Local Similarity 99.5%; Pred. No. 3.3e-204; Matches 912; Conservative 0; Mismatches 2; Indels 3; Gaps 2;
QY	81 AGGCGAGTGTACTACTTCAACCAATCACTAACGCCCACTGGAGCGCCGACGGCCGAGCGG 140
DB	225 AGGCCGAGTGTACTTCAACCAATCACTAACGCCCACTGGAGCGCCGACGGCCGAGCGG 284
QY	141 CAACAGCAGCAGTGGTGGCAAAAACGGCAGGGGAGCCTGCCAGGGTCGGTGTCTGCGCA 200
DB	285 CAACAGCAGCAGTGGTGGCAAAAACGGCAGGGGAGCCTGCCAGGGTCGGTGTCTGCGCA 344
QY	201 CTGTGCTGTGAAGCAGACGACGATCAAGCGCGGCCCTCTGTCTGTGCGGCAGAGAGATCAC 260
DB	345 CTTGCTGCTGAAGCAGACGACGATCAAGCGCGGCCCTCTGTCTGTGCGGCAGAGAGATCAC 404
QY	261 CCGGACCAAGCAGAGGAGGCCCTTGAGCTGTATCAACGGCTATCATCCAGAAGATCAAGTCGGG 320
DB	405 CCGGACCAAGCAGAGGAGGCCCTTGAGCTGTATCAACGGCTATCATCCAGAAGATCAAGTCGGG 464
QY	321 AGAGGAGGACTTTGAGTCTCTGCGCTCACAGTTCAAGCACTGACGCTCAGCCGAGGCCAG 380
DB	465 AGAGGAGGACTTTGAGTCTCTGCGCTCACAGTTCAAGCACTGACGCTCAGCCGAGGCCAG 524
QY	381 GGGAGACTGGGTGCTTTCAGCAGAGGTCAAGTCAGAGCCCAATTTGAGAGCCCTCGGTT 440
DB	525 GGGAGACTGGGTGCTTTCAGCAGAGGTCAAGTCAGAGCCCAATTTGAGAGCCCTCGGTT 584
QY	441 TGGCTCGGACGGGGGAGATGAGCGGGCCCGTGTTCACGGATTCGGCGCATCCACATCAT 500
DB	585 TGGCTCGGACGGGGGAGATGAGCGGGCCCGTGTTCACGGATTCGGCGCATCCACATCAT 644
QY	501 CTTCCGCACTGAGTGAGGGTGGGAGCCGAGGCTTGCCCTCGGGGCAGGCGGGCGGT 560
DB	645 CTTCCGCACTGAGTGAGGGTGGGAGCCGAGGCTTGCCCTCGGGGCAGGCGGGCGGT 704
QY	561 AGGCCGCGCAGCTCCCTCCCTTGCCCGCCAGCCAGTCAGTCGGCCGAACCCCCCACTCCCTGCGCAC 620
DB	705 AGGCCGCGCAGCTCCCTCCCTTGCCCGCCAGCCAGTCAGTCGGCCGAACCCCCCACTCCCTGCGCAC 764
QY	621 GTCAACAGTAATTTATTTGTTCCCAAAATGGGTGGGAGGGGGCCCTTTCAGATTTGGGGGCC 680
DB	765 GTCAACAGTAATTTATTTGTTCCCAAAATGGGTGGGAGGGGGCCCTTTCAGATTTGGGGGCC 824
QY	681 CTGGGGTCCCACTCCCTGTCCATCCCAAGTTGGGGCTGCGACCGCCAGAGATTTCTCCCTTA 740
DB	825 CTGGGGTCCCACTCCCTGTCCATCCCAAGTTGGGGCTGCGACCGCCAGAGATTTCTCCCTTA 884
QY	741 AGGAATTTGACTTTCAGCAGGGGTGGGAGGCTCCACAGACCCAGGCGAGTGTGGTGGGAGGGG 800
DB	885 AGGAATTTGACTTTCAGCAGGGGTGGGAGGCTCCACAGACCCAGGCGAGTGTGGTGGGAGGGG 944
QY	801 TGTTCCAAAGGAGAGGCTTGCTACGAGAGCGCCCGCTGTCCCGCCAGGTGCTGGAGGC 860
DB	945 TGTTCCAAAGGAGAGGCTTGCTACGAGAGCGCCCGCTGTCCCGCCAGGTGCTGGAGGC 1004
QY	861 AGACTTCGAGGCGCGAATTTGTTTCTAGTTTAGGCCACGCTCTCTCTTTCAGTCGCAAAAGGTG 920
DB	1005 AGACTTCGAGGCGCGAATTTGTTTCTAGTTTAGGCCACGCTCTCTCTTTCAGTCGCAAAAGGTG 1064
QY	921 AACACTCATGCGG--CAGCCATGGGCCCTCTGAGGAACCTGTGCAAG-ACCTTTTACCCCC 977
DB	1065 AACACTCATGCGGCCAGCCATGGGCCCTCTGAGCACTGTGACGCCCTTTTCCCCCCC 1124



Db 86 AGTCCGAGTGTACTTCAACACATCACTAACCCAGCCAGTGGAGCGGCCCATCGG 145  
Qy 141 CAACAGCAGCAGTGTGGCAAAACCGGACAGGGGAGCCTGCAGGGTCCGCTGTCTCGCA 200  
Db 146 CAACAGCAGCAGTGTGGCAAAACCGGACAGGGGAGCCTGCAGGGTCCGCTGTCTCGCA 205  
Qy 201 CTTGCTGTGAAGACACAGCAGTCAAGCGGGCCCTGCTCTGCGCGCAGGAGATCAC 260  
Db 206 CTTGCTGTGAAGACACAGCAGTCAAGCGGGCCCTGCTCTGCGCGCAGGAGATCAC 265  
Qy 261 CCGGACCAAGGAGGA-GGCCCTGAGCTCATCAAGCTACATCCAGAGATCAAGTCCG 319  
Db 266 CCGGACCAAGGAGGAGGCGCCCTGAGCTCATCAAGCTACATCCAGAGATCAAGTCCG 325  
Qy 320 GAGAGGAGCACTTGT-AGTCTCTGGCCTCACAGTTTCAGCGACTGCAAGCTCAGCCAGGCC 378  
Db 326 GAGAGGAGCACTTGTAGTCTCTGGCCTCACAGTTTCAGCGACTGCAAGCTCAGCCAGGCC 385  
Qy 379 -AGGGAGACCTGGG-TGCCTTCAGCAGAGGTGATGAGAGCCATTT-----GAAGA 431  
Db 386 AAGGGAGACCTGGTGTGCTTCAGCAGAGGTGATGAGAGCCATTTTGAAGAAGCC 445  
Qy 432 CGCTCGTGTGCTCGGACGCGGAGATGAGCGGGCC---GTCTTCAAGGATTCGG 488  
Db 446 CTCGTCTTGGCTTCGCGAGCGGGGAGATGAGCGGGCCCGTGTTCAGCGATTCGG 505  
Qy 489 CATCCACATCATCT--CGCACTGATGAGGTGGGAGCCAGCGCTGGCTCGCGGGC 546  
Db 506 CATCCACATCATCTTCCGCACTTGTAGTGGGTGGGAGCCAGCGCTGGCTCGCGGGC 565  
Qy 547 AGGGCAGGGCGGTAGCGCGCAGCTCCCGCTTCCCGCAGCGAGTGGCCGAA--CCCC 605  
Db 566 AGGGCAGGGCGGTAGCGCGCAGCTCCCGCTTCCCGCAGCGAGTGGCCGAAACCCCC 625  
Qy 606 CCACTCCCTGCCACCGTCAACAGTATTTATTTTCCACAAATGGCTGGAGGGGGCCCT 665  
Db 626 CCACTCCCTGCCACCGTCAACAGTATTTATTTTCCACAAATGGCTGGAGGGGGCCCT 685  
Qy 666 TCCAGATTGGGGGCTTGGGGTCCCACTCCCTGTCCATCCCAAGTTGGGGTGGCAAG 725  
Db 686 TCCAGATTGGGGGCTTGGGGTCCCACTCCCTGTCCATCCCAAGTTGGGGTGGCAAG 745  
Qy 726 CAGATTCTCCCTTAAGG-NATTGACTTTCAGAGGGGTGGAGGCTCCAGACCCAGGGC 784  
Db 746 CCAATTCTCCCTTAAGGAAATTGACTTTCAGAGGGGTGGAGGCTCCCAAGCCAGGGC 805  
Qy 785 AGTGTGTGGAGGGGTGTTCCAAAGAGAGAGCCCTGCTCAGCAGAGCGCCCGTGTCCC 844  
Db 806 AGTGTGTGGAGGGGTGTTCCAAAGAGAGAGCCCTGCTCAGCAGAGCGCCCGTGTCCC 865  
Qy 845 CCCAGGTGTGGAGGCACTTCGAGGGCGGAATTTTCTAGTTAGGCGAGCTCCTCTG 904  
Db 866 CCCAGGTGTGGAGGCACTTCGAGGGCGGAATTTTCTAGTTAGGCGAGCTCCTCTG 925  
Qy 905 TTAGTGTGCAAGAGTGAACCTCATCGG--CAGCCATGGGCCCTCTGAGCACTGTGCA 962  
Db 926 TTAGTGTGCAAGAGTGAACCTCATCGGCGCCAGCCATGGGCCCTTTGAGCACTGTGCA 985  
Qy 963 G-ACCTTTTCAACCCCAATTAACCCAGAACCA 994  
Db 986 GCACCTTTTCAACCCCAATTAACCCAGATCCA 1018

RESULT 7  
AAS76134  
ID AAS76134 standard; cDNA; 1045 BP.  
XX AAS76134;  
AC AAS76134;  
XX 13-FEB-2002 (first entry)  
DT 13-FEB-2002 (first entry)  
XX DNA encoding novel human diagnostic protein #11938.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US008631.  
XX 31-MAR-2000; 2000US-00540217.  
XX 23-AUG-2000; 2000US-00649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI: 2001-639362/73.  
XX P-PSDB; ABG11947.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX Claim 1; SEQ ID NO 11938; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II) (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 1045 BP; 221 A; 317 C; 324 G; 183 T; 0 U; 0 Other;

Query Match 59.7%; Score 593.8; DB 5; Length 1045;  
Best Local Similarity 85.9%; Pred. No. 2.5e-133;  
Matches 822; Conservative 0; Mismatches 92; Indels 43; Gaps 13;  
Qy 81 AGGCCGAGTGTACTTCAACACATCACTAACCCAGCCAGTGGAGCGGCCCATCGG 140  
Db 87 AGGCCGAGTGTACTTCAACACATCACTAACCCAGCCAGTGGAGCGGCCCATCGG 146  
Qy 141 CAACAGCA-GCAGTGTGGCAAAACCGGACAGGGGAGCCTGCGAG--GGTCCGCTGTCTG 198  
Db 147 CAACAGCATGCAGTGTGGCAAAACCGGACAGGGGAGCCTGCGAGCGGTCCGCTGTCTG 206  
Qy 199 CACTGTGTGTAAGCAACAGCAGTCAAGCCAGTCAAGCCAGTGGAGCGGCCCATCGG 257  
Db 207 CACTGTGTGTAAGCAACAGCAGTCAAGCCAGTCAAGCCAGTGGAGCGGCCCATCGG 266  
Qy 258 CACCCGACCAAGGAGGAGGCGCCCTGGAGCTGATCAAGCGGTACATCCAGAGATCAAGTC 317  
Db 267 CACCCGACCAAGGAGGAGGCGCCCTGGAGCTGATCAAGCGGTACATCCAGAGATCAAGTC 326



318	GGGAGAGGAGGACTTT--GAGTCTCTGGCCCTCA CAGTTTCAGCGACT---- <td>372</td>	372
Qy		
327	GGGAGAGGAGGACTTTGGAGTCTCTGGCCCTCA CAGTTTCAGCGACTCGAAGCATCAGCCAA	386
Db		
373	AAGGCCAGGGGACACTGGGTGCGTT-- --CAGCAGAGGT CAGATGCAAGAGCCATT-- --	426
Qy		
387	AGCCCAAGGGGAGACCTGGGTGCGTTTTCAGCAGAGGT CAGATTGCAAGAGCCATTTTGA	446
Db		
427	--GAAGACGCGCTGGTTTTCGCTCGCGGACGGGGAGATGAGCGGCGCCCGTGT-- --CAGCGA	482
Qy		
447	AGAACGCGCTGGTTTTCGCTTCGCGACGGGGAGATGACGGGCGCGTGTGTTTCACGCAAT	506
Db		
483	TTCCGGGATCCACATCATCTCTCCGCACTGAGTG-- --AGGGTGGGAGGCCACAGGCGTGGCC	539
Qy		
507	TTCCGGGATCCACATCATCTCTTCGCGACTTGAGTGAGGGTTGGGAGGCCACAGGCGTGGCC	566
Db		
540	TCGGGGCAGGCGAGGGCGGCTAGCGCCGCGCCACGCTCCCGCTTCGCCGCCACGACGATGGCGG	599
Qy		
567	TCGGGGCAGGCGAGGGCGGCTAGCGCGGCCACGCTCCCGCTTCGCCGCCACGACGATGGCGG	626
Db		
600	AAACCCCCACTCCCTGCGCACCGTCCACACAGTATTATTGTTCCCCACATATGCTGGGAGGG	659
Qy		
627	AAACCCCCACTCCCTGCGCACCGTCCACACAGTATTATTGTTCCCCACATATGCTGGGAGGG	686
Db		
660	GGCCCTTCCAGATTGGGGCGCCCT-----GGGGTCCCCACTCCCTGTC	701
Qy		
687	GGCCCTTCCAGACTGGGGCGCCCTCCAGACTGGGGCGCCCTGGGGTTCGCCACTCCCTGTC	746
Db		
702	CATCCCAGATTGGGGCTCGGACCGGCAGATTCTCCCTTAAGGAATTGACTTTCAGCAGGG	761
Qy		
747	AGTCCCAGATTGGGGCTCGGACCTCCAGACTCTCCATTAAAGAGTGTGACTTCAGCATGGG	806
Db		
762	TGGAGAGCTCCACAGACCCAGGCGAGTGTGTGGAGGGGTGTTCAAAGAGAAGCCCTGG	821
Qy		
807	AGGAGAGCTCCAGACCCAGGCGAATGTGGTGGGAGGAGTGTTCAGAGAGAGATCTGG	866
Db		
822	TCAGCAGAGCCGCCCGGTGTCCCCCCAGGTTGTGAGGCGAGACTCGAGGGCC-- --GAATTG	879
Qy		
867	TCAGCAGAGCCGCCCGCATGTCTCCACACAGGTCTCGAGGCAGATGCGAGGGCCCGAGTTG	926
Db		
880	TTTCTAGTTAGGCCACGCTCTCTGTTCAGTCGCAAGAGGTGAACACTCATGGG-- --GCAGCC	938
Qy		
927	TTTGTAGTTAGGCCACGCTCTCTGTTCAGTAGCAAAAGCAAACTCGTGGGCCCGAGCT	986
Db		
939	ATGGCCCTCTGAGCAACTGTGAG-- --ACCTTTTACCCCCCAATTAAACCGAGACCA	994
Qy		
987	GTGGCCCTCTGAGCAACTTTCAGCACCCCTTTCACCCCAATTAATCCGAGACCA	1043
Db		

## PRESIDENT 8

RESULI 8  
AAZ33510  
ID AAZ33510 standard; cDNA; 580 BP.  
XX  
XX  
XX AC  
XX AC  
XX  
XX DT 08-DEC-1999 (first entry)  
XX  
XX DE Human prostate cancer-associated EST 34.  
XX  
XX KW Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;  
XX KW cancer; tissue specificity; human; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN DE19811194-A1.  
XX  
XX PD 16-SEP-1999.  
XX  
XX PF 10-MAR-1998; 98DE-01011194.  
XX  
XX PR 10-MAR-1998; 98DE-01011194.  
XX  
XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
XX

[illegible]

## RESULT 9

ACH32387  
ID ACH32387 standard; cDNA; 458 BP.  
XX AC  
XX ACH32387;  
XX  
DT 13-OCT-2003 (first entry)  
XX  
XX Human endothelial cell cDNA #520.  
XX  
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
XX genome mapping; biodiversity; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX US2003073623-A1.  
XX  
XX 17-APR-2003.  
XX  
XX 30-JUL-2001; 2001US-00918995.  
XX  
XX 30-JUL-2001; 2001US-00918995.  
XX  
XX (DRMA/) DRMANAC R T.  
XX (LABA/) LABAT I.  
XX (STAC/) STACHE-CRAIN B.  
XX (DICK/) DICKSON M C.  
XX (JONE/) JONES L W.  
XX  
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
XX  
XX WPI; 2003-615964/58.  
XX  
XX New polynucleotide sequences obtained from various cDNA libraries, useful  
XX as hybridization probes, as oligomers for PCR, for chromosome and gene  
XX mapping, in the recombinant production of protein, or in generating  
XX antisense DNA or RNA.  
XX  
XX Claim 1; SEQ ID NO 19599; 44pp; English.  
XX  
XX The invention relates to an isolated polynucleotide comprising any one of  
XX 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
XX determined by the technique of SBH (sequencing by hybridisation). Also  
XX included is a purified polypeptide comprising a sequence corresponding to  
XX a reading frame of the novel polynucleotide. The nucleic acid sequences  
XX are useful in diagnostics as expressed sequence tags (EST) for  
XX identifying expressed genes or for physical mapping of the human genome,  
XX in forensics, in assessing biodiversity, or in identifying mutations  
XX responsible for genetic disorders and other traits. The nucleotide  
XX sequences are also useful as hybridisation probes, as oligomers for PCR,  
XX for chromosome and gene mapping, in the recombinant production of  
XX protein, or in generating antisense DNA or RNA. The purified polypeptide  
XX is useful for generating antibodies specific for it. The present sequence  
XX is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
XX for this patent did not form part of the printed specification, but was  
XX obtained in electronic format directly from USPTO at  
XX seqdata.uspto.gov/sequence.html?DocID=20030073623  
XX  
XX Sequence 458 BP; 103 A; 147 C; 123 G; 84 T; 0 U; 1 Other;  
XX  
XX Query Match 39.4%; Score 392; DB 8; Length 458;  
XX Best Local Similarity 99.3%; Pred. No. 9.3e-85;  
XX Matches 415; Conservative 0; Mismatches 0; Indels 3; Gaps 2;  
XX  
QY 580 TCCCGCCAGCAGTGGCGGACACCCCACTCCCTGCCACCGTCCACAGTATTATTGT 639  
DB 17 TCCCGCCAGCAGTGGCGGACACCCCACTCCCTGCCACCGTCCACAGTATTATTGT 76  
QY 640 TCCCAATGGTGGGAGGGGCGCTCCAGATTGGGGGCGCTGGGCTCCCACTCCCTG 699  
DB 77 TCCCAATGGTGGGAGGGGCGCTCCAGATTGGGGGCGCTGGGCTCCCACTCCCTG 136  
QY 700 TCCATCCCGAGTGGGCTGGGACCGCCAGATTCTCCCTTAAGGAATTGACATTCAGCAGG 759

DB 137 TCCATCCCGAGTGGGCTGGGACCGCCAGATTCTCCCTTAAGGAATTGACATTCAGCAGG 196  
QY 760 GGTGGAGGCTCCAGACCCAGGGCAGTGTGGTGGAGGGGTCTTCCAAAGAGAGGCCT 819  
DB 197 GGTGGAGGCTCCAGACCCAGGGCAGTGTGGTGGAGGGGTCTTCCAAAGAGAGGCCT 256  
QY 820 GGTCCAGCAGAGCGCGCCCGTGTCCCGCAGGTCTGGAGGCAGACTCGAGGGCCGAATTG 879  
DB 257 GGTCCAGCAGAGCGCGCCCGTGTCTCCCGCAGGTCTGGAGGCAGACTCGAGGGCCGAATTG 316  
QY 880 TTTCTAGTTAGGCGCCAGCTCTCTGTTCAGTGCAGAAAGGTGAACACTCATGCGG--CAGC 937  
DB 317 TTTCTAGTTAGGCGCCAGCTCTCTGTTCAGTGCAGAAAGGTGAACACTCATGCGGCGCAGC 376  
QY 938 CATGGGCGCTCTGAGCAACTGTGTCAG-ACCCTTTACCCCCCAATTAAACCCAGAACCA 994  
DB 377 CATGGGCGCTCTGAGCAACTGTGTCAGCACCCTTTACCCCCCAATTAAACCCAGAACCA 434  
RESULT 10  
ID ABA62612/C  
XX ABA62612 standard; DNA; 559 BP.  
XX AC  
XX ABA62612;  
XX  
DT 01-FEB-2002 (first entry)  
XX  
XX Human foetal liver single exon nucleic acid probe #10917.  
XX  
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX Homo sapiens.  
XX  
XX WO200157277-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000669.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human foetal liver.  
XX  
XX Claim 1; SEQ ID NO 10917; 639pp + Sequence Listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for measuring  
XX human gene expression in a sample derived from human foetal liver. The  
XX single exon nucleic acid probes may be used for predicting, measuring and  
XX displaying gene expression in samples derived from human foetal liver. The  
XX present sequence is a single exon nucleic acid probe of the invention.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 559 BP; 103 A; 176 C; 168 G; 112 T; 0 U; 0 Other;  
XX  
XX Query Match 21.9%; Score 217.6; DB 4; Length 559;  
XX Best Local Similarity 96.1%; Pred. No. 1.1e-42;  
XX Matches 223; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

QY      65 GCATGAGCGCGAGCTCAGGCGAGTGTAATACTTCAACCAATCACTAACGCCAGCCAGT 124
Db      544 GCCTGCCCTCCCTCCAGGCGAGTGTAATACTTCAACCAATCACTAACGCCAGCCAGT 485

QY      125 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAAACGGGCGAGGAGCTGCGCA 184
Db      484 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAAACGGGCGAGGAGCTGCGCA 425

QY      185 GGGTCCGCTGCTCGACCTGCTGTTGAAGCACAGCCAGTCAAGCGGCGCTTCTGCTGGC 244
Db      424 GGGTCCGCTGCTCGACCTGCTGTTGAAGCACAGCCAGTCAAGCGGCGCTTCTGCTGGC 365

QY      245 GGGTCCGCTGCTCGACCTGCTGTTGAAGCACAGCCAGTCAAGCGGCGCTTCTGCTGGC 296
Db      424 GGGTCCGCTGCTCGACCTGCTGTTGAAGCACAGCCAGTCAAGCGGCGCTTCTGCTGGC 365

QY      245 GGCAGGAGAGATCAACCGGACCAAGGAGGAGGCGCTTGGAGCTGATCAACGG 296
Db      364 GGCAGGAGAGATCAACCGGACCAAGGAGGAGGCGCTTGGAGCTGATCAACGG 313

RESULT 11
AAI42606/c
ID      AAI42606 standard; DNA; 559 BP.
XX
AC      AAI42606;
XX
DT      17-OCT-2001 (first entry)
XX
DE      Probe #11292 used to measure gene expression in human placenta sample.
XX
KW      Probe; microarray; human; placenta; antenatal diagnosis;
KW      genetic disorder; ss.
XX
OS      Homo sapiens.
XX
PN      WO200157272-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US000663.
XX
PR      04-FEB-2000; 2000US-0180312P.
PR      26-MAY-2000; 2000US-0207456P.
PR      30-JUN-2000; 2000US-00608408.
PR      03-AUG-2000; 2000US-00632366.
PR      21-SEP-2000; 2000US-0234687P.
PR      27-SEP-2000; 2000US-0236359P.
PR      04-OCT-2000; 2000GB-00024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-48897/53.
XX
PT      Human genome-derived single exon nucleic acid probes useful for analyzing
PT      gene expression in human placenta.
XX
PS      Claim 25; SEQ ID NO 11292; 654pp; English.
XX
CC      The present invention relates to single exon nucleic acid probes (SENP).
CC      The present sequence is one such probe. The probes are useful for
CC      producing a microarray for predicting, measuring and displaying gene
CC      expression in samples derived from human placenta. The probes are useful
CC      for antenatal diagnosis of human genetic disorders
XX
SQ      Sequence 559 BP; 103 A; 176 C; 168 G; 112 T; 0 U; 0 Other;
XX
Query Match      21.9%; Score 217.6; DB 4; Length 559;
Best Local Similarity 96.1%; Pred. No. 1.1e-42;
Matches 223; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      65 GCATGAGCGCGAGCTCAGGCGAGTGTAATACTTCAACCAATCACTAACGCCAGCCAGT 124
Db      544 GCCTGCCCTCCCTCCAGGCGAGTGTAATACTTCAACCAATCACTAACGCCAGCCAGT 485

```

```

QY      125 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAAACGGGCGAGGAGCTGCGCA 184
Db      484 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAAACGGGCGAGGAGCTGCGCA 425

QY      185 GGGTCCGCTGCTCGACCTGCTGTTGAAGCACAGCCAGTCAAGCGGCGCTTCTGCTGGC 244
Db      424 GGGTCCGCTGCTCGACCTGCTGTTGAAGCACAGCCAGTCAAGCGGCGCTTCTGCTGGC 365

QY      245 GGCAGGAGAGATCAACCGGACCAAGGAGGAGGCGCTTGGAGCTGATCAACGG 296
Db      364 GGCAGGAGAGATCAACCGGACCAAGGAGGAGGCGCTTGGAGCTGATCAACGG 313

RESULT 12
AAK36818/c
ID      AAK36818 standard; DNA; 559 BP.
XX
AC      AAK36818;
XX
DT      06-NOV-2001 (first entry)
XX
DE      Human bone marrow expressed single exon probe SEQ ID NO: 11375.
XX
KW      Human; bone marrow expressed exon; gene expression analysis; probe;
KW      microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS      Homo sapiens.
XX
PN      WO200157276-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US000668.
XX
PR      04-FEB-2000; 2000US-0180312P.
PR      26-MAY-2000; 2000US-0207456P.
PR      30-JUN-2000; 2000US-00608408.
PR      03-AUG-2000; 2000US-00632366.
PR      21-SEP-2000; 2000US-0234687P.
PR      27-SEP-2000; 2000US-0236359P.
PR      04-OCT-2000; 2000GB-00024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-488900/53.
XX
PT      Human genome-derived single exon nucleic acid probes useful for analyzing
PT      gene expression in human bone marrow.
XX
PS      Example 4; SEQ ID NO 11375; 658pp + Sequence Listing; English.
XX
CC      The present invention provides a number of single exon nucleic acid
CC      probes which are derived from genomic sequences expressed in the human
CC      bone marrow. They can be used to measure gene expression in bone marrow
CC      samples, which may enable the improved diagnosis and treatment of cancers
CC      such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC      the probes of the invention
XX
SQ      Sequence 559 BP; 103 A; 176 C; 168 G; 112 T; 0 U; 0 Other;
XX
Query Match      21.9%; Score 217.6; DB 4; Length 559;
Best Local Similarity 96.1%; Pred. No. 1.1e-42;
Matches 223; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      65 GCATGAGCGCGAGCTCAGGCGAGTGTAATACTTCAACCAATCACTAACGCCAGCCAGT 124
Db      544 GCCTGCCCTCCCTCCAGGCGAGTGTAATACTTCAACCAATCACTAACGCCAGCCAGT 485

QY      125 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAAACGGGCGAGGAGCTGCGCA 184
Db      484 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAAACGGGCGAGGAGCTGCGCA 425

```

```

QY 185 GGGTCCGCTGCTCGACCTGCTGTAAGCAACAGCCAGTCAAGCGGCGCCCTGCTCTGGC 244
D 424 GGGTCCGCTGCTCGACCTGCTGTAAGCAACAGCCAGTCAAGCGGCGCCCTGCTCTGGC 365
QY 245 GGCAGGAGAGATCACCCGGACCAAGGAGGAGGCGCCCTGGAGCTGATCAACGG 296
D 364 GGCAGGAGAGATCACCCGGACCAAGGAGGAGGCGCCCTGGAGCTGATCAACGG 313

RESULT 13
AAK10973/c
ID AAK10973 standard; DNA; 559 BP.
XX
AC AAK10973;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 10964.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.
XX
OS Homo sapiens.
XX
FN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-0004263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
XX
PS Example 4; SEQ ID NO 10964; 650pp + Sequence Listing; English.
XX
The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention
XX
SQ Sequence 559 BP; 103 A; 176 C; 168 G; 112 T; 0 U; 0 Other;
Query Match 21.9%; Score 217.6; DB 4; Length 559;
Best Local Similarity 96.1%; Pred. No. 1.1e-42;
Matches 223; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 65 GCATGAGCGCAGCTCAGGCGGAGTGACTTCAACACATCACTAAGCGCCAGT 124
D 544 GCTTCCCTCCCTCCAGGCGGAGTGACTTCAACACATCACTAAGCGCCAGT 485
QY 125 GGGAGCGGCCAGCGCAACAGCAGAGTGTGGGCAAAACGGGAGGAGCTGGCA 184
D 484 GGGAGCGGCCAGCGCAACAGCAGAGTGTGGGCAAAACGGGAGGAGCTGGCA 425

```

```

QY 185 GGGTCCGCTGCTCGACCTGCTGTAAGCAACAGCCAGTCAAGCGGCGCCCTGCTCTGGC 244
D 424 GGGTCCGCTGCTCGACCTGCTGTAAGCAACAGCCAGTCAAGCGGCGCCCTGCTCTGGC 365
QY 245 GGCAGGAGAGATCACCCGGACCAAGGAGGAGGCGCCCTGGAGCTGATCAACGG 296
D 364 GGCAGGAGAGATCACCCGGACCAAGGAGGAGGCGCCCTGGAGCTGATCAACGG 313

RESULT 14
ABS10823/c
ID ABS10823 standard; DNA; 559 BP.
XX
AC ABS10823;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe from lung SEQ ID NO 10814.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
FN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US000665.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-0004263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.
XX
PS Claim 1; SEQ ID NO 10814; 634pp; English.
XX
The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of probes
CC ; the novel set of probes which hybridise at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)
CC algorithmically predicting at least one exon from genomic sequences of
CC the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,

```

CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karagenen syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and biliary membrane disease. The  
CC present sequence is a single exon probe of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 559 BP; 103 A; 176 C; 168 G; 112 T; 0 U; 0 Other;  
Query Match 21.9%; Score 217.6; DB 6; Length 559;  
Best Local Similarity 96.1%; Pred. No. 1.1e-42;  
Matches 223; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 65 GCATGAGCGGCGAGCTCAGCGCGAGTGTACTTCAACACATCACTAAGCGGCGAGCT 124  
DB 544 GCGTGCCTCCCTCCAGCGCGAGTGTACTTCAACACATCACTAAGCGGCGAGCT 485  
QY 125 GGGAGCGGCGCGCGCAACAGCAGCAGTGGTGGCAAAACCGGGAGCGCTGGCCA 184  
DB 484 GGGAGCGGCGCGCGCAACAGCAGCAGTGGTGGCAAAACCGGGAGCGCTGGCCA 425  
QY 185 GGGTCCGCTGCTCGACCTGCTGGTGAAGCAGCCAGTCAAGCGGCGCTGCTGGG 244  
DB 424 GGGTCCGCTGCTCGACCTGCTGGTGAAGCAGCCAGTCAAGCGGCGCTGCTGGG 365  
QY 245 GGCAGGAGAGATCACC CGGACCAAGGAGGAGGAGGCGCTGGAGCTGATCAACGG 296  
DB 364 GGCAGGAGAGATCACC CGGACCAAGGAGGAGGAGGCGCTGGAGCTGATCAACGG 313

RESULT 15  
ABA75131/c  
ID ABA75131 standard; DNA; 213 BP.  
XX ABA75131;  
AC ABA75131;  
DT 01-FEB-2002 (first entry)  
XX  
XX Human foetal liver single exon nucleic acid probe #23436.  
DE Human foetal liver single exon nucleic acid probe #23436.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX Homo sapiens.  
OS  
XX W0200157277-A2.  
PN  
XX  
XX 09-AUG-2001.  
FD  
XX 30-JAN-2001; 2001WO-US000669.  
PF  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human fetal liver.

XX Claim 4; SEQ ID NO 23436; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring  
XX human gene expression in a sample derived from human foetal liver. The  
XX single exon nucleic acid probes may be used for predicting, measuring and  
XX displaying gene expression in samples derived from human fetal liver. The  
XX present sequence is a single exon nucleic acid probe of the invention.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 213 BP; 25 A; 72 C; 67 G; 49 T; 0 U; 0 Other;

Query Match 21.4%; Score 213; DB 4; Length 213;  
Best Local Similarity 100.0%; Pred. No. 1.1e-41;  
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 83 GCGGAGTGTACTTCAACACATCACTAAGCGGCGAGTGGGAGCGGCGGCGCA 142  
DB 213 GCGGAGTGTACTTCAACACATCACTAAGCGGCGAGTGGGAGCGGCGGCGCA 154  
QY 143 ACAGCAGCAGTGGTGGCAAAACCGGGAGCGGCGGAGCTGCTGCTGCTGCTG 202  
DB 153 ACAGCAGCAGTGGTGGCAAAACCGGGAGCGGCGGAGCTGCTGCTGCTGCTG 94  
QY 203 TGCTGGTGAAGCAGCAGCCAGTCAAGCGGCGCTGCTGCTGCGGCGGAGGAGTCA 262  
DB 93 TGCTGGTGAAGCAGCAGCCAGTCAAGCGGCGCTGCTGCTGCGGCGGAGGAGTCA 34  
QY 263 GGACCAAGGAGGAGGCGCTGGAGCTGATCAACG 295  
DB 33 GGACCAAGGAGGAGGCGCTGGAGCTGATCAACG 1

Search completed: June 28, 2004, 16:36:42

Job time : 550.575 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:25 ; Search time 113.078 Seconds  
(without alignments)  
4878.251 Million cell updates/sec

Title: U49070  
Perfect score: 994  
Sequence: 1 TGTGGCCAGCACCTCGAGS.....CCCAATTAAACCCAGAACCA 994

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	994	100.0	1014	2	US-09-066-074-1
2	994	100.0	1014	2	US-08-555-912A-1
3	994	100.0	1014	4	US-09-275-900-1
4	65.4	6.6	295	4	US-09-313-294A-5899
5	53.6	5.4	989	4	US-09-507-242-1
6	51.8	5.2	32998	4	US-09-408-020-1
7	49	4.9	42432	4	US-09-408-020-2
8	48.2	4.8	279	4	US-09-408-020-67
9	47.8	4.8	2097	4	US-09-252-991A-13635
10	47.8	4.8	2331	4	US-09-252-991A-13809
11	45.8	4.6	7218	1	US-08-232-463-14
12	45	4.5	1608	4	US-09-252-991A-11839
13	45	4.5	2640	4	US-09-252-991A-11547
14	45	4.5	2934	4	US-09-252-991A-11690
15	44.2	4.4	430	4	US-09-621-976-16656
16	44.2	4.4	1266	4	US-09-252-991A-2521
17	44.2	4.4	1377	4	US-09-252-991A-2310
18	44.2	4.4	1428	4	US-09-252-991A-2606
19	43	4.3	1260	4	US-09-252-991A-14282
20	43	4.3	2175	4	US-09-252-991A-14313
21	43	4.3	2244	4	US-09-252-991A-14391
22	42	4.2	446	4	US-09-621-976-17380
23	42	4.2	480	4	US-09-621-976-17379
24	42	4.2	497	4	US-09-621-976-17378
25	42	4.2	605	3	US-09-208-804-2
26	42	4.2	605	3	US-08-801-743-2
27	42	4.2	605	4	US-09-016-434-270

Sequence 13686, A  
Sequence 13459, A  
Sequence 13887, A  
Sequence 1673, Ap  
Sequence 725, App  
Sequence 725, App  
Sequence 26, Appli  
Sequence 7, Appli  
Sequence 1, Appli  
Sequence 859, App  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1342, Ap  
Sequence 3451, Ap  
Sequence 3440, Ap  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 4, Appli

## ALIGNMENTS

RESULT 1  
US-09-066-074-1  
; Sequence 1, Application US/09066074  
; Patent No. 5952467  
; GENERAL INFORMATION:  
; APPLICANT: Hunter et al., Tony  
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/066,074  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/555,912  
; FILING DATE: 13-NOV-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Liesa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07251/011001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1014 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 25..513  
US-09-066-074-1

Query Match 100.0%; Score 994; DB 2; Length 1014;  
Best Local Similarity 100.0%; Pred. No. 6.2e-224;  
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGGCCAGCACCTCGAGGAGAGAGTGGCGGAGGAGAGCTGCCCGCGCTGGAG 60

RESULT 2  
US-08-555-912A-1  
; Sequence 1, Application US/08555912A  
; Patent No. 5972697  
; GENERAL INFORMATION:

Query Match	100.0%; Score 994; DB 2; Length 1014;
Best Local Similarity	100.0%; Pred. No. 6.2e-224;
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 TCGTGGCCAGCACCTCGAGGGAAGATGGCGGACGAGGAGAACTGCGCGCGCGCTGGGAG 60
DB	1 TGTGTGCCAGCACCTCGAGGGAAGATGGCGGACGAGGAGAACTGCGCGCGCGCTGGGAG 60
QY	61 AAGCGCATGAGCGCGAGCTTCAGGCGGAGTGTACTTCAAACCATCTACTAAGCCAGC 120
DB	61 AAGCGCATGAGCGCGAGCTCAGGCGGAGTGTACTTCAAACCATCTACTAAGCCAGC 120
QY	121 CAGTGGGAGCGGCCGAGCGGCAACAGCACAGTGTGGCAAAACGGGACGGGGAGCCT 180
DB	121 CAGTGGGAGCGGCCGAGCGGCAACAGCACAGTGTGGCAAAACGGGACGGGGAGCCT 180
QY	181 GCCAGGCTCGCTGCTCGCACCTGTGTGTGAAGCACAGCCAGTCAGCGCGGCCCTCGTCC 240
DB	181 GCCAGGCTCGCTGCTCGCACCTGTGTGTGAAGCACAGCCAGTCAGCGCGGCCCTCGTCC 240
QY	241 TGGCGGCAGGAGAAGATCACCCGGACCAAGGAGGAGGCCCTGAGTCTGATCAACGGGTAC 300
DB	241 TGGCGGCAGGAGAAGATCACCCGGACCAAGGAGGAGGCCCTGAGTCTGATCAACGGGTAC 300
QY	301 ATCCAGAAGATCAAGTTCGGGAGGAGGACTTTGAGTCTCTGCGCCTCACAGTTCAGGCAC 360
DB	301 ATCCAGAAGATCAAGTTCGGGAGGAGGACTTTGAGTCTCTGCGCCTCACAGTTCAGGCAC 360
QY	361 TGCAGCTCAGCCCAAGSCCAGGCGAGACCTTGGGTGCTTTCAGCAGAGGTCAAGATGCAGAAG 420
DB	361 TGCAGCTCAGCCCAAGSCCAGGCGAGACCTTGGGTGCTTTCAGCAGAGGTCAAGATGCAGAAG 420
QY	421 CCATTTGAAGACGCCTCGTTTGGCGCTTCGGAACGGGGAGATGAGCGGGGCCGTGTTCAAG 480
DB	421 CCATTTGAAGACGCCTCGTTTGGCGCTTCGGAACGGGGAGATGAGCGGGGCCGTGTTCAAG 480



```
QY 481 GATTCGGGCAATCCATCATCTCCGCACTGAGTGAGGAGGAGCCGAGCCCTGGCT 540
Db 481 GATTCGGGCAATCCATCATCTCCGCACTGAGTGAGGAGGAGCCGAGCCCTGGCT 540
QY 541 CGGGCAGGGCAGGGCGGTGAGCGCGCCAGCTCCCTTGGCCCGCAGCCAGCTGGCGA 600
Db 541 CGGGCAGGGCAGGGCGGTGAGCGCGCCAGCTCCCTTGGCCCGCAGCCAGCTGGCGA 600
QY 601 ACCCCCACTCCCTCCGCACTGAGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
Db 601 ACCCCCACTCCCTCCGCACTGAGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
QY 661 GCGCTTCCAGATGGGGGCGCTGGGGTCCCGCTCCCTGTCATCCCTGTCATCCCTGTC 720
Db 661 GCGCTTCCAGATGGGGGCGCTGGGGTCCCGCTCCCTGTCATCCCTGTCATCCCTGTC 720
QY 721 GACCGCAGATTCCTCCCTTAAGGAATGACTTCAGCAGGGGTGGGAGGCTCCCGACCCA 780
Db 721 GACCGCAGATTCCTCCCTTAAGGAATGACTTCAGCAGGGGTGGGAGGCTCCCGACCCA 780
QY 781 GGGCAGTGTTGGGAGGGGTGTTCCAAAGAGAGGCTGGTTCAGCAGAGCCCGCCGCTG 840
Db 781 GGGCAGTGTTGGGAGGGGTGTTCCAAAGAGAGGCTGGTTCAGCAGAGCCCGCCGCTG 840
QY 841 TCCCCCAAGTGTGGAGGAGCAGCTCAGAGGCGGAAATGTTCTAGTTAGGCGACGCTCC 900
Db 841 TCCCCCAAGTGTGGAGGAGCAGCTCAGAGGCGGAAATGTTCTAGTTAGGCGACGCTCC 900
QY 901 TCTGTTGAGTCCGAAAGGTGAACACTCATGGGCGAGCATGGGCGCTCTCAGCAACTGTG 960
Db 901 TCTGTTGAGTCCGAAAGGTGAACACTCATGGGCGAGCATGGGCGCTCTCAGCAACTGTG 960
QY 961 CAGACCCCTTTCACCCCAATTAACCCAGAACCA 994
Db 961 CAGACCCCTTTCACCCCAATTAACCCAGAACCA 994
```

RESULT 3

```
US-09-275-900-1
; Sequence 1, Application US/09275900
; Patent No. 6596848
; GENERAL INFORMATION:
; APPLICANT: Hunter et al., Tony
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/275,900
; FILING DATE: 24-Mar-1999
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/555,912
; FILING DATE: 13-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07251/011001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
```

```
;
; LENGTH: 1014 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..513
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-275-900-1
```

```
Query Match 100.0%; Score 994; DB 4; Length 1014;
Best Local Similarity 100.0%; Pred. No. 6.2e-224;
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGTGTCGCAAGCACTTCGAGGGAAGATGGCGGAGGAGAGCTGCCCGCGCTGGAG 60
Db 1 TGTGTCGCAAGCACTTCGAGGGAAGATGGCGGAGGAGAGCTGCCCGCGCTGGAG 60
QY 61 AAGCGCATGAGCGCGAGCTCAGGCCGAGTGTACTTCAACACATCACTAACGCCAGC 120
Db 61 AAGCGCATGAGCGCGAGCTCAGGCCGAGTGTACTTCAACACATCACTAACGCCAGC 120
QY 121 CAGTGGAGCGGCCCAAGCGGCAACAGCAGTGTGTGGCAAAAACGCGGAGGAGCCT 180
Db 121 CAGTGGAGCGGCCCAAGCGGCAACAGCAGTGTGTGGCAAAAACGCGGAGGAGCCT 180
QY 181 GCCAGGGTCGGTGTGCGACCTGCTGAGCAGCAGCCAGTCACGGCGGCGCTCGTCC 240
Db 181 GCCAGGGTCGGTGTGCGACCTGCTGAGCAGCAGCCAGTCACGGCGGCGCTCGTCC 240
QY 241 TGGCGGCGAGGAGAGATCACCCGAGCAAGAGGAGGAGCCCTGGAGCTGATCAACGGGTAC 300
Db 241 TGGCGGCGAGGAGAGATCACCCGAGCAAGAGGAGGAGCCCTGGAGCTGATCAACGGGTAC 300
QY 301 ATCCAGAGATCAAGTGGGAGAGGAGCACTTTGAGTCTCTGGCCTCACAGTTCAGGCAC 360
Db 301 ATCCAGAGATCAAGTGGGAGAGGAGCACTTTGAGTCTCTGGCCTCACAGTTCAGGCAC 360
QY 361 TGCAGCTCAGCCAGGCGGAGGAGAGCTGGTGGCTTCAGCAGAGCTCAGATCAGAGAG 420
Db 361 TGCAGCTCAGCCAGGCGGAGGAGAGCTGGTGGCTTCAGCAGAGCTCAGATCAGAGAG 420
QY 421 CCATTTGAAGACGCTCGTTTGGCTCGGAGCGGGGAGATGAGCGGCGCTGTTCAAG 480
Db 421 CCATTTGAAGACGCTCGTTTGGCTCGGAGCGGGGAGATGAGCGGCGCTGTTCAAG 480
QY 481 GATTCCGGCATCCATCATCTCCGCACTGAGTGAGGAGGAGGAGGAGGAGGAGGAGG 540
Db 481 GATTCCGGCATCCATCATCTCCGCACTGAGTGAGGAGGAGGAGGAGGAGGAGGAGG 540
QY 541 CGGGCAGGGCAGGGCGGCTAGGCGGCGCTAGGCGGCGAGCTCCCCCTTGCCTCCAGTGGCCGA 600
Db 541 CGGGCAGGGCAGGGCGGCTAGGCGGCGCTAGGCGGCGAGCTCCCCCTTGCCTCCAGTGGCCGA 600
QY 601 ACCCCCACTCCCTGCCACCGCTCACAGATATTTATTTGTTCCCAATFGCTGGAGGGG 660
Db 601 ACCCCCACTCCCTGCCACCGCTCACAGATATTTATTTGTTCCCAATFGCTGGAGGGG 660
QY 661 GCGCTTCCAGATGGGGGCGCTGGGGTCCCGCTCCCTGTCATCCCTGTCATCCCTGTC 720
Db 661 GCGCTTCCAGATGGGGGCGCTGGGGTCCCGCTCCCTGTCATCCCTGTCATCCCTGTC 720
QY 721 GACCGCAGATTCCTCCCTTAAGGAATGACTTCAGCAGGGGTGGGAGGCTCCCGACCCA 780
Db 721 GACCGCAGATTCCTCCCTTAAGGAATGACTTCAGCAGGGGTGGGAGGCTCCCGACCCA 780
QY 781 GGGCAGTGTTGGGAGGGGTGTTCCAAAGAGAGGCTGGTTCAGCAGAGCCCGCCGCTG 840
Db 781 GGGCAGTGTTGGGAGGGGTGTTCCAAAGAGAGGCTGGTTCAGCAGAGCCCGCCGCTG 840
QY 841 TCCCGCCAGGTCGTGGAGGAGCAGCTCAGAGGCGGAAATGTTCTAGTTAGGCGACGCTCC 900
Db 841 TCCCGCCAGGTCGTGGAGGAGCAGCTCAGAGGCGGAAATGTTCTAGTTAGGCGACGCTCC 900
```



901	TTCTGTTGAGTGC	AAGTGAAC	TCTATCGGC	CAGCCATGGG	CCCTCTGAG	CAACTGTG	960
QY							
901	TTCTGTTGAGTGC	AAGTGAAC	TCTATCGGC	CAGCCATGGG	CCCTCTGAG	CAACTGTG	960
DB							
961	CAGACCCCTTTT	CACCCCCCAATT	TAACCCAG	ACCA			994
QY							
961	CHGACCCCTTTT	CACCCCCCAATT	TAACCCAG	ACCA			994
DB							

## RESULT 4

```

US-09-313-294A-5899
; Sequence 5899, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 5899
; LENGTH: 295
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Inbyte ID No. 6476212 700350866H1
; NAME/KEY: unsure
; LOCATION: 244, 247, 252-255, 260-261, 276, 291
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5899

```

	Query Match	Score 65.4;	DB 4;	Length 295;
	Best Local Similarity 66.9%;	Pred. No. 1.5e-06;		
	Matches 93;	Conservative 0;	Mismatches 46;	Indels 0; Gaps 0
Qy	385	GACCTGGGTGCTCTCAGCAGAGGTCAGATGCAAGCCATTGGAAGAGCCCTCGTTTCGG	444	
Db	1	GACTTAGTACTTTGGAGGCGGAGATGAGAAACCTTCGAGGAGCCACATATGCC	60	
Qy	445	CTGCGGAGCGGGGAGATGAGCGGGCCGTGTTTCACGGATTCGGGATCCACATCATCCTC	504	
Db	61	CTCAAGTTTGGTAGCTCAGCGACATCGTGGACACTGACAGTGGGGTTTCACATCATCCTG	120	
Qy	505	CGCACTGAGTGAGGTGGG	523	
Db	121	CGCACTGCTGAGGAGGGG	139	

## RESIST 5

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

; TYPE: DNA
; ORGANISM: Candida albicans
; US-09-507-242-1

Query Match      5.4%; Score 53.6; DB 4; Length 989;
Best Local Similarity 48.4%; Pred. No. 0.0011;
Matches 180; Conservative 0; Mismatches 189; Indels 3; Gaps 1;

QY      148 AGCAGTGGTGGCAAAACCGCGACGGGGAGCTCGCCAGAGTCCGCTCTCGCACTCGTGTG 207
Db      418 AACAAATGGTTACAAAGCCACTTGTGAATGAGATGGCCAGGTAGAGTTTCTCATTTGTTG 477

QY      208 GTCAAGCACACGCGAGTCAAGCGCGGCCCTCTCTCTGGCGG--CAGGAGAAAGATCACCGGG 264
Db      478 ATCAAGAACAAATCAATCAAGAAAAACCCAAAGTCTTGGAAAGTCCCAGAGTGGTATAGTAGA 537

QY      265 ACCAAGAGAGAGGCCCTGGAGCTGATCAACGGCTACATCCAGAGAGATCAAGTCGGGAGAG 324
Db      538 ACTAGAGACAAATCTATACAGATATTGAAGAAACATTTGGAAAGATATTGAGTGGTGAG 597

QY      325 GAGGACTTTTGAGTCTCTGGGCTCACAGTTTCAGCGACTCGAGCTCAGCCAAAGCCAGGGGA 384
Db      598 GTTAAACTAAGTGAATTTGGCAAAATACCGAAAGTGATTGCACTCACATGACAGAGTGGT 657

QY      385 GACCTGGTGCCTTCACGACAGGTCAGATGCAGAAGCCATTGGAAGAGCCTCGTTTTCGG 444
Db      658 GATTTAGGGTTTTTTTAGCAAGGACAAATGCAACCACTTCGAAGAAAGCCGCAATTCAT 717

QY      445 CTGCGGACGGGGAGATGAGCGGGCCCGTGTTCAGGATTCGGGCATCCACATCATCTCTC 504
Db      718 TTGATGTTGGAGAGTCAGTAAACATAATTGAACCAATAGTGGTGTCTCATATCTCTCCA 777

QY      505 CGCACTGAGTGA 516
Db      778 AGAACAGGATAA 789

```

## RESULT 6

```

US-09-408-020-1/c
; Sequence 1, Application US/09408020
; Patent No. 6632937
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCRP.002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 32998
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7604) ... (8908)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8961) ... (9767)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10545) ... (10922)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13944) ... (14612)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (18638) ... (20149)
; FEATURE:

```

u49070.rni

Tue Jun 29 08:37:22 2004

```

; NAME/KEY: CDS
; LOCATION: (20554)...(20955)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (20956)...(21834)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25151)...(26377)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27535)...(28002)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (28065)...(29483)
US-09-408-020-1

Query Match
Best Local Similarity 5.2%; Score 51.8; DB 4; Length 32998;
Matches 146; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 311 TCAAGTCGGAGAGGAGGACTTTGAGTCTCTGGCTCTCAGTTCAGGACTGCAGCTCAG 370
Db 20361 TCAAGCGGGGGAAGTTTGAAAGCTGGCAAGAGCTCTCGATAGACGGGGGCGAGCG 20302

QY 371 CCAAGGCCAGGGAGAGCTGGTGCTTCAGCAGAGTTCAGATGCAGAGCCATTGGAAG 430
Db 20301 CAAAGAGGACGCGCAGCTTGGGCTACTTTGGCAGGGGCAAGATGTAAGCCGTTTGAGG 20242

QY 431 ACGCTCGTTTGGCTGGGACGGGGAGATGAGCGGGCCGCTGTTACGGGATTCGGCA 490
Db 20241 ATGCGCGGTTCCGCTTCGAGTGGAGGATTCGAGCGGTAAATCCGAGTTTGGCT 20182

QY 491 TCCACATCATCTCCGCACTCAGTGAGGTTGGGAGCCAGCCAGCTGGCTCGGGGCAAGG 550
Db 20181 ACCAGCTGATAAGGCTTGGGATAAGCCGCTAGAAATAGCCCTCGAGCCCTTTT 20122

QY 551 CAGGCGGCTAGGCGGCGAGCTCCCTTGGCCCGCAGCCAGTGGCGGAAACCCCTACT 610
Db 20121 GGGGCTGCTTTGGGCGAGGGGCCCCCGCTGCAAGCGACTTGTTCCTGTCCTCCCATG 20062

QY 611 CCC 613
Db 20061 CCC 20059
```

```

RESULT 7
US-09-408-020-2/c
; Sequence 2, Application US/09408020
; Patent No. 6632937
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP 002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 42432
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(10421)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10625)...(11434)
; FEATURE:
; NAME/KEY: CDS
```

```

; LOCATION: (11478)...(13046)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13046)...(14620)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23558)...(24862)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (24913)...(25728)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (26504)...(26881)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29655)...(30491)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34559)...(36067)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37002)...(37403)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37404)...(38282)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (39454)...(40572)
US-09-408-020-2

Query Match
Best Local Similarity 4.9%; Score 49; DB 4; Length 42432;
Matches 94; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 352 TTCAGCGACTCAGTCCAGCAAGCCAGGCGGAGACCTGGGTGCTTCAGCAGAGGTCCAG 411
Db 36228 TCCATGGACGGGGGAGTGCAGAGGAGGACGCGAGTCTCGGCTACTTTGGCAGGGGAAAG 36169

QY 412 ATGCAAGAGCCATTGGAAGACGCTTCGTTGCGCTGCGGAGCGGGGAGATGAGCGGGCCC 471
Db 36168 ATGTAAGAGCCGTTTGAAGAGCGCGGTTCCGGCTGCAGATAGGCGAGATCTCGGAACCC 36109

QY 472 GTGTTACGAGATTCGGGATCCACATCATCTCCGCACTGAGTGGGT 520
Db 36108 ATAAAGTCAGAGTTTGGCTATCACGTGATAAAGCGCTCGGCTAAAGT 36060
```

```

RESULT 8
US-09-408-020-67
; Sequence 67, Application US/09408020
; Patent No. 6632937
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP 002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 279
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(279)
US-09-408-020-67

Query Match
4.8%; Score 48.2; DB 4; Length 279;
```

Best Local Similarity 52.8%; Pred. No. 0.016; Mismatches 93; Indels 0; Gaps 0;  
Matches 104; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
QY 311 TCAAGTCGGAGAGGAGGACTTTCAGTCTCTGGCTCAGCTTCAGGACTGACGCTCAG 370  
Db 74 TCAAGCGGGGGAAGAGTTTGAAGCTGCGCAAGGAGCTCTCGATACGCGGGCAGCG 133  
QY 371 CCAAGCGGAGGAGAGACTGGGTGCTTCAGAGAGGTTCAGTCAAGAGCCATTGAG 430  
Db 134 CAAAGAGGAGCGGAGCTTGGGCTACTTTGGCAGGGCAAGATGTTAAAGCGGTTGAGG 193  
QY 431 AGCCTCTGTTGCGTGGCGGAGAGTACGCGGCGGTGTTACGAGTTCGGCA 490  
Db 194 ATGCGCGGTTCGCTGCGAGTAGGAGGATTCGAGCGGTAAATCCGAGTTTGCT 253  
QY 491 TCCACATCATCTCCGC 507  
Db 254 ACCAGTGATAAGCGC 270

RESULT 9  
US-09-252-991A-13635  
; Sequence 13635, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13635  
; LENGTH: 2097  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13635

Query Match 4.8%; Score 47.8; DB 4; Length 2097;  
Best Local Similarity 49.8%; Pred. No. 0.03;  
Matches 149; Conservative 0; Mismatches 147; Indels 3; Gaps 1;  
QY 206 TGGTGAAGCACAGCCAGTCAAGCGGCGCTCTGCTGGCGGAGGAGAGATCACCGGA 265  
Db 1004 TCGCCAACTCTCCGAGCAGCGGATCGCGCCACATCTCTGATCGAGGTGAACGCAAGG 1063  
QY 266 CCAAGGAGGAGGCGCTGAGTCAAGCGGTACATCCAGAGATCAAGTCCGGAGAGG 325  
Db 1064 TCGGCGAGCAGCAGGCGCAAGCGGAGATCAGCAGAGATCAAGGCTCGCTGGCCAGGGCG 1123  
QY 326 AGGACTTGTAGTCTCTGCTCTCAGTTCAGC---GACTGAGCTCAGCCAGGCGCAGG 382  
Db 1124 AGGATTTGCGCGCTGGCCAGAGGTTCTCCAGAGTATCGGCTCGCGCCACCGCG 1183  
QY 383 GAGACCTGGGTGCTTCAGCAGAGGTTCAGTCAAGAGCCATTGAGAGCGCTCGTTG 442  
Db 1184 GCGACCTGGGTACGCGCGCGCGGTGTACGAGCCCGCGCTTCGAGGAGCGCTGTATG 1243  
QY 443 CGCTCGGAGCGGGAGATGAGCGGCGCGGTTCACGGATTCGGGATCCACATCATC 501  
Db 1244 CGCTGAAGCAAGGTAGGTATCCGCGCGCGGTGAAGACTCCGTAAGCTACCACTGATC 1302

RESULT 10  
US-09-252-991A-13809/C  
; Sequence 13809, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13809  
; LENGTH: 2331  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13809

Query Match 4.8%; Score 47.8; DB 4; Length 2331;  
Best Local Similarity 49.8%; Pred. No. 0.031;  
Matches 149; Conservative 0; Mismatches 147; Indels 3; Gaps 1;  
QY 206 TGGTGAAGCACAGCCAGTCAAGCGGCGCTCTGCTGGCGGAGGAGATCACCGGA 265  
Db 1505 TCGCCAACTCTCCGAGCAGCGGATCGCGCCACATCTGATCGAGGTGAACGCAAGG 1446  
QY 266 CCAAGGAGGAGGCGCTGAGTCAAGCGGTACATCCAGAGATCAAGTCCGGAGAGG 325  
Db 1445 TCGGCGAGCAGCAGGCGCAAGCGAGATCGAGAGATCAAGGCTCGCTGGCCAGGGCG 1386  
QY 326 AGGACTTGTAGTCTCTGCTCTCAGTTCAGC---GACTGAGCTCAGCCAGGCGCAGG 382  
Db 1385 AGGATTTGCGCGCTGGCCAGGAGTTCCTCCAGAGTATCGGCTCGCGCCACCGCG 1326  
QY 383 GAGACCTGGGTGCTTCAGCAGAGGTTCAGATCAGAGCCATTGAGAGCGCTCGTTG 442  
Db 1325 GCGACTGGGTACGCGCGCGCGGTGTACGACCCCGCGTTCGAGGAGGCGCTGTATG 1266  
QY 443 CGCTCGGAGCGGGAGATGAGCGGCGCGGTTCACGGATTCGGGATCCCGCATCCATCATC 501  
Db 1265 CGCTGAAGCAAGGTAGGTATCCGCGCGCGGTGAAGACTCCGTAAGCTACCACTGATC 1207

RESULT 11  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFELINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgt-F1s  
US-08-232-463-14

	Query Match	4.6%;	Score 45.8;	DB 1;	Length 7218;	
	Best Local Similarity	4.9%;	Pred. No. 0.12;			
	Matches	20;	Conservative 216;	Mismatches 173;	Indels 0;	Gaps 0;
QY	17	GAGGGAAGTGGCGGACGAGAGAAGCTGCGCCCGCTGGGAGAGCAGCATGACGCCA	76			
Db	- 1433	RRR	1374			
QY	77	GCTCAGCCGCAAGTGTTACTTCCAACAATCACTAAACGCCAGCCAGTCGGAGCGGCCCA	136			
Db	1373	RRR	1314			
QY	137	GGCGCAACAGCACAGCTGTTGGCAAAAACGGGACAGGGGAGCTGCAGGGTCGCTGCT	196			
Db	1313	RRR	1254			
QY	197	CGCACCTGTGTTGAACACAGCAGCTACGGGGGCCCTCGTCTGGCGCAGGAGAAGA	256			
Db	1253	RRR	1194			
QY	257	TCAACCCGGAACAAGGAGGACCCTGGAGCTGATCAACGGCTACATCCAGAAGTCAAGT	316			
Db	1193	RRR	1134			
QY	317	CGGAGAGGAGGACTTTTGAGTCTCTGGCCTCACAGTTTCAGCGACTGCAGCTCAGCCAAAG	376			
Db	1133	RRR	1074			
QY	377	CCAGGGGAGACCTGGTGCCCTTCAGCAGAGGTCAGATCGAGAACCCATT	425			
Db	1073	RRRRRRATCCCAAGTCCCTCGACTCGACCTCGACCAAGCTCGAAATAATT	1025			

```

RESULT 12
US-09-252-991A-11839
; Sequence 11839, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11839
; LENGTH: 1608
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-11839
Query Match          4.5%; Score 45; DB 4; Length 1608;

```

	Best Local Similarity	47.7%; Pred. No. 0.13;			
	Matches 132; Conservative	0; Mismatches 145; Indels	0; Gaps	0;	
QY	52	GGCTTGGGAGAAGCCATGAGCCCGAGCTCAGGCCGAGTGTA	TACTTTCAACCATCACT	111	
Ddb	953	GTCTTGGGCATGTGTCGCGCCAGCAGCGCACCTGGCGGACTACGA	CCCCGGGACC	1012	
QY	112	AACGCGAGCCMGTGGGHAGCGGCCAAGCAGCAGCAGTAGTGTG	GCAAAAACGGGCAG	171	
Ddb	1013	AAGSGCAGCATTTGGCGAGCTCGCACCGGCCCTGTGTTCAATCCC	GAGGCCGCGCAGCGG	1072	
QY	172	GGGAGAGCTGCGCAGGCTCCGTGCTCGCACCTGTGTTGAAGCA	CAGCCAGTCAAGCGGG	231	
Ddb	1073	GTCAGCATTCGCCGTTGGCTACTTCGGGTACGTGATGCGT	CGACTGGCCAAGCTT	1132	
QY	232	CCCTCGTCTCTGGCGCAGGAGAATACCCGGACCAAGGAGGAGCC	CTGGAGCTGATC	291	
Ddb	1133	GGCGTCGCCCTCCGCGAGGACGCGCTGCAGTTACGTACCTG	ACCACCCAGGTGTCC	1192	
QY	292	RACGGCTACATCCAGAAGATCAAGTCGGCAGGAGG	328		
Ddb	1193	TACGAGAGGCCAACATGCTCAACACCGAGACCGG	1229		

RESULT 13  
US-09-252-991A-11547/c  
; Sequence 11547, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 11547  
; LENGTH: 2840  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11547

Query Match	4.5%;	Score 45;	DB 4;	Length 2640;
Best Local Similarity	47.7%;	Pred. No. 0.14;		
Matches 132;	Conservative 0;	Mismatches 145;	Indels 0;	Gaps 0;
QY	52	GGCTGGGAGAAAGCGCATGAGCCGACGTGAGCCGAGTGTACTCTTCAACACATCACT	111	
DB	1818	GTCTGGGACATGCTGGTGGCGGACGAGCGCCACTGGGCGACTACGACCCGGGGACC	1759	
QY	112	AACGCCAGCCAGTGGGAGCGGCCAGCGGCAACAGACGAGTGGTGGCAAAAACGGGCGAG	171	
DB	1758	AAGGGCAGCATTGGCGAGCTGCGCACCGGCGCTGGTTCAATCCGAGGCGCGGCACGCGG	1699	
QY	172	GGGAGCGCTGCAGGGTCCGCTGCTGCACCTGCTGGTGAAGCACAGACGAGTCAACGGCGG	231	
DB	1698	GTCAGCATTCGCGGGTGGGCTACTCCGGCTACGTGATGGCTCGGACCTGGCCACAGCTT	1639	
QY	232	CCCTCGTCTTGGCGGCGAGGAGAAGATCACCCGGACCAAGGAGGAGGCCCTTGGAGCTGATC	291	
DB	1638	GGCGTGGCCCTTGGCGAGGACCAGCGCTGCAGTTTCAGCTACCTGACCACCCAGGTGTCC	1579	
QY	292	AACGCTACATCCAGAAAGATCAAGTCCGGAGAGGAGG	328	
DB	1578	TACGACGAGCGCAACATGCTCAACACCGAGAACCCAGG	1542	
RESULT 14				
US-09-252-991A-11690				

; Sequence 11690, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 11690  
; LENGTH: 2934  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11690

Query Match 4.5%; Score 45; DB 4; Length 2934;  
Best Local Similarity 47.7%; Pred. No. 0.15;  
Matches 132; Conservative 0; Mismatches 145; Indels 0; Gaps 0;  
QY 52 GGCCTGGAGAGCGCATGAGCGGAGCTCAGCGGAGTGTACTTCAACACATCACT 111  
DB 1174 GTCTGGGACATGCTGGTCCGCCGACGAGCGCCACCTGGGCGACTACGACCGGGGACC 1233  
QY 112 AACGCCAGCCAGTGGGAGCGGCCACGCGGCAACAGCAGCAGTGGTGGCAAAAACGGGCGAG 171  
DB 1234 AAGGGCAGCATTTGGCGAGCTGGCGACCGCGCGCTGGTTCAATCCCGAGGCGGGCAGCGG 1293  
QY 172 GGGGAGCCTGCCAGGCTCGCTGCTGCACTCTGCTGTAAGCAGACAGCCAGTCAACGGCGG 231  
DB 1294 GTCAAGCATTCGCGGCTGGGCTACTCGGCTACGTGATGCGCTCGGACTGGCCAAAGCTT 1353  
QY 232 CCCTGCTGCTGGCGGAGGAGAGATCACCGGACCAAGGAGGAGCGCCCTGGAGCTGATC 291  
DB 1354 GGGCTGCGCCCTGGCGAGGACGACGCGCTGATGCTAGTACCTGACCCAGGTGTCC 1413  
QY 292 AACGGCTTACATCCAGAGATCAAGTGGGAGAGGAGG 328  
DB 1414 TAGGACGACGCCAATGCTCAACACCGAGAACGAGG 1450

RESULT 15  
US-09-621-976-16656  
; Sequence 16656, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 16656  
; LENGTH: 430  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-16656

Query Match 4.4%; Score 44.2; DB 4; Length 430;  
Best Local Similarity 18.9%; Pred. No. 0.15;  
Matches 49; Conservative 101; Mismatches 109; Indels 0; Gaps 0;  
QY 156 TGGCAAAACGGGAGGGGAGCCTGCCAGGCTCGCTGCGACCTGCTGGTGAAGCA 215  
DB 31 TGCCMCMWRRSCCYMKSGGRRSCGRAMCCYYKKXGSCRAMCCCCYTYKSCSS 90

QY 216 CAGCCAGTCACGGCGCCCTCTGCTCTGGCGCAGGAGAAGATCACCCGACCAAGAGGGA 275  
DB 91 YKGSYTTKRAMWKRSCYTYRRRMYYYWRSYMMRSMWKGSCCCGSGGSCYKKKKKGK 150  
QY 276 GGCCTGGAGTGTATCAACGGCTACATCCAGAAGATCAAGTCGGGAGAGGAGCTTTGA 335  
DB 151 KGSCCMRSYWCCYYKRAARMWMMKGGSCMYYTKRMWRRMCCCMRRRRSMRRRRMCMWKG 210  
QY 336 GTCTCTGGCCTTCACAGTTTCAGCGACTGCGAGCTCAGCCAAAGGCCAGGGGAGACCTGGGTGC 395  
DB 211 SYTCYKSSSMCMARWKRARGKKRMCCYTKGGGRWWRKYCMRKKGRACCTGTTCAC 270  
QY 396 CTTCAGCAGAGGTGATG 414  
DB 271 CTGCCGTGCTGTCGAGAG 289  
Search completed: June 28, 2004, 18:04:34  
Job time : 115.078 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 18:36:26 ; Search time 389 Seconds  
(without alignments)  
118.296 Million cell updates/sec

```
Title: AAC50492
Perfect score: 852
Sequence: 1 MADEKLPQGWKRMSSSG.....GEMSGPVFTDSGIHILRTE 163
```

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length:	Maximum DB seq length:
2000000000	2000000000

Post-processing: Minimum Match 0%

100% Processing. Maximum Match 0%  
Maximum Match 100%

## Listing first 45 summaries

Database : Published Applications AA:\*\*

- 1: /cgn2\_6/prodata/2/pubpaa/US07\_PUBCOMB.pcp.\*  
2: /cgn2\_6/prodata/2/pubpaa/PCT\_NEW\_PUB.pcp.\*  
3: /cgn2\_6/prodata/2/pubpaa/US06\_NEW\_PUB.pcp.\*  
4: /cgn2\_6/prodata/2/pubpaa/US06\_PUBCOMB.pcp.\*  
5: /cgn2\_6/prodata/2/pubpaa/US07\_NEW\_PUB.pcp.\*  
6: /cgn2\_6/prodata/2/pubpaa/PCTUS\_PUBCOMB.pcp.\*  
7: /cgn2\_6/prodata/2/pubpaa/US08\_NEW\_PUB.pcp.\*  
8: /cgn2\_6/prodata/2/pubpaa/US09\_PUBCOMB.pcp.\*  
9: /cgn2\_6/prodata/2/pubpaa/US09A\_PUBCOMB.pcp.\*  
10: /cgn2\_6/prodata/2/pubpaa/US09B\_PUBCOMB.pcp.\*  
11: /cgn2\_6/prodata/2/pubpaa/US09C\_PUBCOMB.pcp.\*  
12: /cgn2\_6/prodata/2/pubpaa/US09\_NEW\_PUB.pcp.\*  
13: /cgn2\_6/prodata/2/pubpaa/US10A\_PUBCOMB.pcp.\*  
14: /cgn2\_6/prodata/2/pubpaa/US10B\_PUBCOMB.pcp.\*  
15: /cgn2\_6/prodata/2/pubpaa/US10C\_PUBCOMB.pcp.\*  
16: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pcp.\*  
17: /cgn2\_6/prodata/2/pubpaa/US60\_NEW\_PUB.pcp.\*  
18: /cgn2\_6/prodata/2/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	852	100.0	163	16	US-10-687-361-2		Sequence 2, Appli
2	665	78.1	145	15	US-10-108-260A-4441		Sequence 4441, Ap
3	634	62.7	105	16	US-10-687-361-16		Sequence 16, Appl
4	437.5	51.3	178	14	US-10-128-714-3443		Sequence 3443, Ap
5	434.5	51.0	178	14	US-10-128-714-8443		Sequence 8443, Ap
6	406	47.7	87	14	US-10-263-828-116		Sequence 116, App
7	373	43.8	70	9	US-09-864-761-4607A		Sequence 4607A, A
8	369.5	43.4	177	14	US-10-342-555-2		Sequence 2, Appli
9	369.5	43.4	177	14	US-10-032-595-7374		Sequence 7374, Ap
10	361	42.4	209	15	US-10-320-797-3033		Sequence 3033, Ap
11	318	37.3	126	12	US-10-424-599-185102		Sequence 185102,
12	314	36.9	126	12	US-10-424-599-185101		Sequence 185101,
13	306.5	36.0	188	12	US-10-425-114-41877		Sequence 41877, A
14	306.5	36.0	204	12	US-10-425-114-65823		Sequence 65823, A
15	243	28.5	107	16	US-10-687-361-17		Sequence 17, Appl

## ATTIGNEMENTS

## RESULT 1

```

US-10-687-361-2
; Sequence 2, Application US/10687361
; Publication NO. US20040101896A1
; GENERAL INFORMATION:
; APPLICANT: Hunter, Tony
; APPLICANT: Lu, Kun Ping
; TITLE OF INVENTION: NIMA INTERACTION
; FILE REFERENCE: 66671-078
; CURRENT APPLICATION NUMBER: US/10/10-687-361-2
; CURRENT FILING DATE: 2003-10-15
; PRIOR APPLICATION NUMBER: US 10/61-87361
; PRIOR FILING DATE: 2003-07-08
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Vers 1.0
; SEQ ID NO 2
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-687-361-2

```

Query Match	100.0%;	Score 852;	DB 16;	Length 163;
Best Local Similarity	100.0%;	Pred. No. 2.2e-80;		

1 MADEEXT.PPGWEKRMSSSGRIYVFNHTTNA SOWERPSCNSSSCCKNGCCEPDPVBCSEI. 60

1 MA DEEXT.PPGWEKTRMSPSSGRIYVENHTTNASOWEPBSCNSSSSCCVNCORRPAIBPCSHI 60

QY 61 LVKHSQSRPPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120

61 LVKHSQSRPSSWRQEKITRTKEEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120

QY 121 DLGAFSRGQMQPFEDASFALRTGEMSGPVFTDSGIHILRTE 163

db 121 DLGAFSRGQMCKPFEDASFALRTGEMSGPVFTDSGHIILRTE 163

```

RESULT 2
US-10-108-260A-4441
; Sequence 4441, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4441
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4441

Query Match      78.1%; Score 665; DB 15; Length 145;
Best Local Similarity 98.4%; Pred. No. 5.2e-61;
Matches 126; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MADEKLPGWKMSRSGRVYFNHITNASQWERPSGSGGNGGQGEAPARVRCSHL 60
DB 1 MADEKLPGWKMSRSGRVYFNHITNASQWERPSGSGGNGGQGEAPARVRCSHL 60

QY 61 LVKHSOSRRPSSWRQEKITRTKEALELINGYTIQIKSGEEDFESLASQFSDCSSAKARG 120
DB 61 LVKHSOSRRPSSWRQEKITRTKEALELINGYTIQIKSGEEDFESLASQFSDCSSAKARG 120

QY 121 DLGAFSRG 128
DB 121 DLGAFSRG 128

RESULT 3
US-10-687-361-16
; Sequence 16, Application US/10687361
; Publication No. US20040101896A1
; GENERAL INFORMATION:
; APPLICANT: Hunter, Tony
; APPLICANT: Lu, Kun Ping
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; FILE REFERENCE: 666711-078
; CURRENT APPLICATION NUMBER: US/10/687,361
; CURRENT FILING DATE: 2003-10-15
; PRIOR APPLICATION NUMBER: US 10/616,410
; PRIOR FILING DATE: 2003-07-08
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-687-361-16

Query Match      62.7%; Score 534; DB 16; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.4e-47;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 HLLVKHSOSRRPSSWRQEKITRTKEALELINGYTIQIKSGEEDFESLASQFSDCSSAKA 118
DB 1 HLLVKHSOSRRPSSWRQEKITRTKEALELINGYTIQIKSGEEDFESLASQFSDCSSAKA 60

QY 119 RGDLGAFSRGQMKPFEDASFALRTEGMSGPVFTDSTGHIILRTE 163
DB 61 RGDLGAFSRGQMKPFEDASFALRTEGMSGPVFTDSTGHIILRTE 105

RESULT 4
US-10-128-714-3443
; Sequence 3443, Application US/10128714
; Publication No. US20030119013A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wendi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 3443
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3443

Query Match      51.3%; Score 437.5; DB 14; Length 178;
Best Local Similarity 51.5%; Pred. No. 3e-37;
Matches 88; Conservative 26; Mismatches 40; Indels 17; Gaps 3;

QY 7 LPGWKMSRSGRVYFNHITNASQWERPSG-----NSSSGK---NGOG 50
DB 9 LPAGWEVRHSNKNLPYFNPTKESRWEPSPGDTTELKVTYVNHVSAPAGRPDGTAG 68

QY 51 EPARYVCSHLLVKHSOSRRPSSWRQEKITRTKEALELINGYTIQIKSGEEDFESLASQF 110
DB 69 E-GKIRCSEHLLKHDRSRPSSWRREAITRSKEAEIILRGHEQIRSRSEVSLGDIATSE 127

QY 111 SDSSAKARGDLGAFSRGQMKPFEDASFALRTEGMSGPVFTDSTGHIILR 161
DB 128 SDSSAKRGDLGFFRGEMQKFEFADAFALQPGVSGIVETASGVHLLIER 178

RESULT 5
US-10-128-714-8443
; Sequence 8443, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wendi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31

```



```
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8443
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8443

Query Match      51.0%; Score 434.5; DB 14; Length 178;
Best Local Similarity 50.6%; Pred. No. 6.1e-37;
Matches 88; Conservative 26; Mismatches 43; Indels 17; Gaps 3;

QY 4 BEKLPPEWEMSRSSGRVYFNHITNASQWERPSG-----NSSSGGK---N 47
DB 6 ETGLPAGWEVRHNSKNLPYFNSTKESRWPESGDTTELLKVMYMHYSAPAGRDPGT 65
QY 48 GQGPAPVRCSHLLVKGHSQSRPSSRWQEKITRKEEALELINGYIQIKSGEEDFSLA 107
DB 66 AQGE-GKIRCSHLLIKHRDSRPPSSWREAEITRSKEAEITLKGHEQIRSGEVSGLDIA 124
QY 108 SQFSDCSAKARGDLGAFSGROMQKPPEDASFAIARTGEMSGPVTDGSIHILR 161
DB 125 VSESDCSSARKKDCGFFRGEMQKEFEADAFAALQPGQGVIVETASGVHLIER 178

RESULT 6
US-10-263-828-116
; Sequence 116, Application US/10263828
; Publication No. US20030138905A1
; GENERAL INFORMATION:
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Glenn, Matthew R.
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; TITLE OF INVENTION: Compositions isolated from bovine
; TITLE OF INVENTION: mammary gland and methods for their use.
; FILE REFERENCE: 11000.1044ULCON
; CURRENT APPLICATION NUMBER: US/10/263,828
; CURRENT FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 136
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 116
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Bovine
US-10-263-828-116

Query Match      47.7%; Score 406; DB 14; Length 87;
Best Local Similarity 96.1%; Pred. No. 2.2e-34;
Matches 74; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MADEKLPPEWEMSRSSGRVYFNHITNASQWERPSGSSGGKNGGQGPAPVRCSHL 60
DB 1 MADEKLPPEWEMSRSSGRVYFNHITNASQWERPSGSSGGKNGGQGPAPVRCSHL 60
QY 61 LVKHSQSRPPSSRWQEK 77
DB 61 LVKHSQSRPPSSRWQEK 77

RESULT 7
US-09-864-761-46074
; Sequence 46074, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
```

```
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46074
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008752.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.4
; OTHER INFORMATION: EST HUMAN HIT: BF314489.1, EVALUATE 2.00e-28
; OTHER INFORMATION: SWISSPROT HIT: Q13526, EVALUATE 2.00e-29
US-09-864-761-46074

Query Match      43.8%; Score 373; DB 9; Length 70;
Best Local Similarity 100.0%; Pred. No. 4.5e-31;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 RYYFNHITNASQWERPSGSSGGKNGGQGPAPVRCSHLLVKGHSQSRPPSSRWQEKLTR 80
DB 1 RYYFNHITNASQWERPSGSSGGKNGGQGPAPVRCSHLLVKGHSQSRPPSSRWQEKLTR 80
QY 81 TKEEALELIN 90
DB 61 TKEEALELIN 70

RESULT 8
US-10-342-555-2
; Sequence 2, Application US/10342555
; Publication No. US20030143615A1
; GENERAL INFORMATION:
; APPLICANT: HANES, Steven D.
```

APPLICANT: DEVASAHAYAM, Gina  
APPLICANT: CHATURVEDI, Vishnu  
TITLE OF INVENTION: CAESI: A CANDIDA ALBICANS GENE, METHODS FOR MAKING AND  
TITLE OF INVENTION: USING, AND TARGETING IT OR ITS EXPRESSION PRODUCTS FOR  
TITLE OF INVENTION: ANTIFUNGAL APPLICATIONS  
FILE REFERENCE: 454311-2200.1  
CURRENT APPLICATION NUMBER: US/10/342,555  
CURRENT FILING DATE: 2003-01-15  
PRIOR APPLICATION NUMBER: US/09/507,242  
PRIOR FILING DATE: 2000-02-18  
PRIOR APPLICATION NUMBER: 60/121,246  
PRIOR FILING DATE: 1999-02-23  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 177  
TYPE: PRT  
ORGANISM: Candida albicans  
US-10-342-555-2

Query Match 43.4%; Score 369.5; DB 14; Length 177;  
Best Local Similarity 43.2%; Pred. No. 3.5e-30;  
Matches 73; Conservative 34; Mismatches 49; Indels 13; Gaps 2;

QY 7 LPPGWEKMSRSSGRVYFNHITNASQWERPSGN-----SSSGKNGQGEPA 54

Db 8 LPPNWTIRVSRSHNKYFQINQSTNESSWDPPYGTGKEVLNAYIAKFNKNGYKPLVNEGQ 67

QY 55 VRCSHLLVKHSQSRPPSWRQ-EKTRTKKEALELINGYQIKKSGDEDFESLASQFSDC 113

Db 68 VRVSHLLIKNNQSRKPKSWKSPDGISRTDSEIQILKXHLERILSGEVLSELANTESDC 127

QY 114 SSARKAGDLGAFSRGQMKPEDASFAIRTGEMSGPVFTDSGHIILRT 162

Db 128 SSHDRGGDLGFFSKGQMPPEEAFNLHVGEVNSIIETNSGVHILQRT 176

## RESULT 9

US-10-032-585-7374  
Sequence 7374, Application US/10032585  
Publication No. US20030180953A1  
GENERAL INFORMATION:  
APPLICANT: Terry, Roemer D.  
APPLICANT: Bo, Jiang  
APPLICANT: Charles, Boone  
APPLICANT: Howard, Bussey  
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
FILE REFERENCE: 10182-005-999  
CURRENT APPLICATION NUMBER: US/10/032,585  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 8000  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 7374  
LENGTH: 177  
TYPE: PRT  
ORGANISM: Candida albicans  
US-10-032-585-7374

Query Match 43.4%; Score 369.5; DB 14; Length 177;  
Best Local Similarity 43.2%; Pred. No. 3.5e-30;  
Matches 73; Conservative 34; Mismatches 49; Indels 13; Gaps 2;

QY 7 LPPGWEKMSRSSGRVYFNHITNASQWERPSGN-----SSSGKNGQGEPA 54

Db 8 LPPNWTIRVSRSHNKYFQINQSTNESSWDPPYGTGKEVLNAYIAKFNKNGYKPLVNEGQ 67

QY 55 VRCSHLLVKHSQSRPPSWRQ-EKTRTKKEALELINGYQIKKSGDEDFESLASQFSDC 113

Db 68 VRVSHLLIKNNQSRKPKSWKSPDGISRTDSEIQILKXHLERILSGEVLSELANTESDC 127

QY 114 SSARKAGDLGAFSRGQMKPEDASFAIRTGEMSGPVFTDSGHIILRT 162

Db 128 SSHDRGGDLGFFSKGQMPPEEAFNLHVGEVNSIIETNSGVHILQRT 176

## RESULT 10

US-10-320-797-3033  
Sequence 3033, Application US/10320797  
Publication No. US20040014955A1  
GENERAL INFORMATION:  
APPLICANT: Eroshkin, Alexey M.  
APPLICANT: Zamudio, Carlos  
TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND  
TITLE OF INVENTION: METHODS OF USE  
FILE REFERENCE: 10182-021-999  
CURRENT APPLICATION NUMBER: US/10/320,797  
CURRENT FILING DATE: 2002-12-16  
PRIOR APPLICATION NUMBER: 60/341,261  
PRIOR FILING DATE: 2001-12-17  
NUMBER OF SEQ ID NOS: 3361  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3033  
LENGTH: 209  
TYPE: PRT  
ORGANISM: Cryptococcus neoformans  
US-10-320-797-3033

Query Match 42.4%; Score 361; DB 15; Length 209;  
Best Local Similarity 45.6%; Pred. No. 3.3e-29;  
Matches 82; Conservative 23; Mismatches 43; Indels 32; Gaps 5;

QY 10 GWEKMSRSSGRVYFNHITNASQWERPSGNSS-----SSGKNGQ 49

Db 34 GWEIRFSNRQIFYPYNSERSISTWEPPELSAEQIQLPGAAYKMNVLQAPAGKKEG 93

QY 50 GEPARVRCSHLLVKHSQSRPPSWR--QEKITTKKEALELIN---GYIQIKSGE--ED 102

Db 94 -----VRASHILAKHAGSRPPASWRNVKTTTSDAQAIIEQHITAYLQSLPPADLPKE 148

QY 103 FESLASQFSDCSSAKARGDLGAFSRGQMKPEDASFAIRTGEMSGPVFTDSGHIILRT 162

Db 149 FAXIASTEDCSSARKGDLGWFGRGQMKPFEDATFNPVGQSLGIVKTDGHIHLRT 208

## RESULT 11

US-10-424-599-185102  
Sequence 185102, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 185102  
LENGTH: 126  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_138161C.1.pep  
US-10-424-599-185102

Query Match 37.3%; Score 318; DB 12; Length 126;  
Best Local Similarity 52.8%; Pred. No. 5e-25;  
Matches 67; Conservative 18; Mismatches 36; Indels 6; Gaps 2;

QY 41 SSSGKNGQGEPAVRVRCSHLLVKHSQSRPPSSNRQEK----ITRTKEALELINGYIQI 96

Db 2 SSSSSKGGGG--IEVRASHILITHEGSRKASWADPEGRITKSTTRENAYLSQKALRDDI 59

QY 97 KSGEEDFESLASQFSDCSSAKARGDLGAFSRGQMKPFEDASFAIRTGEMSGPVFTDSGI 156

Db 60 VSGKASFDIASRFSDCSSAKRGDGLGPPRGOMQKPFEEATFALKVGEISDIVDTSKV 119  
157 HIIRTE 163  
120 HIIRTK 126

## RESULT 12

US-10-424-599-185101  
; Sequence 185101, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 185101  
; LENGTH: 126  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_138160C.1.pap  
US-10-424-599-185101

Query Match 36.9%; Score 314; DB 12; Length 126;  
Best Local Similarity 52.0%; Pred. No. 1.3e-24;  
Matches 66; Conservative 19; Mismatches 36; Indels 6; Gaps 2;

QY 41 SSSGKNGGGEAPRVRCSHLLVKHSQSRPPSSWRQK-----ITRTKEALELINGYIQI 96  
DB 2 SSSSKGGGG--TEVRASHILIKHEGSRKASWKQDPGEIIRKSTFENAVSQIKALRDI 59  
QY 97 KSGEEDFESLQFSDCSSAKARGDLGAFSRGOMQKPFEDASFAIRTGMSGPFVTDGSI 156  
DB 60 VSGKASFDIASRFSDCSSAKRGDGLGPPRGOMQKPFEEATFALKVGEISDIVDTSKV 119  
QY 157 HIIRTE 163  
DB 120 HIIRTK 126

## RESULT 13

US-10-425-114-41877  
; Sequence 41877, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53133)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 41877  
; LENGTH: 188  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3136-013-F12\_FLI.pap  
US-10-425-114-41877

Query Match 36.0%; Score 306.5; DB 12; Length 188;  
Best Local Similarity 46.5%; Pred. No. 1.3e-23;  
Matches 66; Conservative 28; Mismatches 33; Indels 15; Gaps 4;  
QY 35 ERPSGNSSSGK-----NGGGEPA-----RVRCSHLLVKHSQSRPPSSWRQK-----ITR 80  
DB 47 EHPS-SSSAGSRDRHHHGRPPSSADEKVRASHILIKHEGSRKASWRDPGVASAT 105  
QY 81 TKEALELINGYIQIKSGEEDFESLQFSDCSSAKARGDLGAFSRGOMQKPFEDASFA 140  
DB 106 TRDDAADLARALRQDIVSGDREFEDVAAENSDCSSAKRGDGLGSGRGOMQKPFKAAPA 165  
QY 141 LRTGMSGPFVFTDSGIHILRT 162  
DB 166 LKVGESIDVVDTESGVHIIRKT 187

## RESULT 14

US-10-425-114-65823  
; Sequence 65823, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53133)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 65823  
; LENGTH: 204  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3354-123-A10\_FLI.pap  
US-10-425-114-65823

Query Match 36.0%; Score 306.5; DB 12; Length 204;  
Best Local Similarity 46.5%; Pred. No. 1.5e-23;  
Matches 66; Conservative 28; Mismatches 33; Indels 15; Gaps 4;

QY 35 ERPSGNSSSGK-----NGGGEPA-----RVRCSHLLVKHSQSRPPSSWRQK-----ITR 80  
DB 63 EHPS-SSSAGSRDRHHHGRPPSSADEKVRASHILIKHEGSRKASWRDPGVASAT 121  
QY 81 TKEALELINGYIQIKSGEEDFESLQFSDCSSAKARGDLGAFSRGOMQKPFEDASFA 140  
DB 122 TRDDAADLARALRQDIVSGDREFEDVAAENSDCSSAKRGDGLGSGRGOMQKPFKAAPA 181  
QY 141 LRTGMSGPFVFTDSGIHILRT 162  
DB 182 LKVGESIDVVDTESGVHIIRKT 203

## RESULT 15

US-10-687-361-17  
; Sequence 17, Application US/10687361  
; Publication No. US20040101896A1  
; GENERAL INFORMATION:  
; APPLICANT: Hunter, Tony  
; APPLICANT: Lu, Kun Ping  
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS  
; FILE REFERENCE: 66671-078  
; CURRENT APPLICATION NUMBER: US/10/687,361  
; CURRENT FILING DATE: 2003-10-15  
; PRIOR APPLICATION NUMBER: US 10/616,410  
; PRIOR FILING DATE: 2003-07-08  
; NUMBER OF SEQ ID NOS: 22

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Yeast ESS1
US-10-687-361-17
```

```
Query Match      28.5%; Score 243; DB 16; Length 107;
Best Local Similarity 48.6%; Pred. No. 2.5e-17;
Matches 51; Conservative 18; Mismatches 34; Indels 2; Gaps 1;
```

```
Qy      59 HLLVKHSQSRPSSRWROEKIIRTKEEALELINGYIQKI--KSGEEDFESLASQFSDCSSA 116
```

```
Db      1 HTLIKHKDSRPASHRSENITISKQDATDELKTLITRLDDDSKTNSFEALAKERSDCSSY 60
```

```
Qy      117 KARGDLGAFSGQMKPFEDASFALRTGEMSGPVTDSGIHILR 161
```

```
Db      61 KRGGDLGWFGRGEMQPSFEDAAFLKVGVEVDIVESGGGVHVIKR 105
```

```
Search completed: June 28, 2004, 18:47:15
Job time : 390 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:24 ; Search time 259.925 seconds  
(without alignments)  
7747.025 Million cell updates/sec

Title: AAC50492

Perfect score: 474

Sequence: 1 ATGGCTGCTACGAGCAAC.....TTCGTGGAGAGCTGCTTAG 474

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseqn\_29Jan04.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002s.\*
- 7: Geneseqn2003as.\*
- 8: Geneseqn2003bs.\*
- 9: Geneseqn2003cs.\*
- 10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	474	100.0	474	3 AAC50492	Aac50492 Arabidops
2	338	71.3	507	3 AAC50504	Aac50504 Arabidops
3	318.6	67.2	531	3 AAC39615	Aac39615 Arabidops
4	301.2	63.5	621	3 AAC39654	Aac39654 Arabidops
5	213.6	45.1	459	3 AAC38179	Aac38179 Arabidops
6	208.6	44.0	411	3 AAC47703	Aac47703 Arabidops
7	208.6	44.0	643	3 AAC51199	Aac51199 Arabidops
8	208.6	44.0	674	3 AAC51182	Aac51182 Arabidops
9	208.6	44.0	677	3 AAC32694	Aac32694 Arabidops
10	207.6	43.8	1005	3 AAC47929	Aac47929 Arabidops
11	207	43.7	646	3 AAC33712	Aac33712 Arabidops
12	205.2	43.3	499	9 AD881685	Ad881685 Arabidops
13	200.8	42.4	699	6 ABK31041	Abk31041 Plant dwa
14	199	42.0	714	3 AAC51198	Aac51198 Arabidops
15	199	42.0	717	3 AAC34378	Aac34378 Arabidops
16	199	42.0	729	6 ABK31048	Abk31048 Plant dwa
17	196.6	41.5	404	7 ABX46988	Abx46988 Bovine SS
18	195.8	41.3	474	3 AAC41338	Aac41338 Zea mays
19	195.8	41.3	492	3 AAC35867	Aac35867 Zea mays
20	195.2	41.2	585	6 ABK31036	Abk31036 Plant dwa
21	194.2	41.0	447	3 AAC51749	Aac51749 Zea mays
22	193.4	40.8	487	3 AAC47648	Aac47648 Arabidops
23	193.4	40.8	669	6 ABK31042	Abk31042 Plant dwa

24	193.4	40.8	682	6 ABK30665	Abk30665 Plant dwa
25	193.4	40.8	682	6 ABK49479	Abk49479 Sequence
26	193.4	40.8	796	6 ABK30897	Abk30897 Plant dwa
27	192.4	40.6	972	3 AAC45109	Aac45109 Arabidops
28	192	40.5	360	6 ABQ85651	Abq85651 Arabidops
29	191.8	40.5	603	6 ABQ66368	Abq66368 Arabidops
30	191.8	40.5	640	3 AAC41085	Aac41085 Arabidops
31	191	40.3	450	6 ABK31037	Abk31037 Plant dwa
32	191	40.3	576	8 ACL22562	ACL22562 DNA clone
33	191	40.3	626	8 ACL22561	ACL22561 DNA clone
34	191	40.3	638	8 ACL22574	ACL22574 DNA clone
35	191	40.3	640	8 ACL22573	ACL22573 DNA clone
36	191	40.3	646	8 ACL22563	ACL22563 DNA clone
37	191	40.3	652	8 ACL22566	ACL22566 DNA clone
38	191	40.3	663	8 ACL22576	ACL22576 DNA clone
39	191	40.3	668	8 ABK31028	Abk31028 Plant dwa
40	191	40.3	672	8 ACL22565	ACL22565 DNA clone
41	191	40.3	678	8 ACL22564	ACL22564 DNA clone
42	191	40.3	688	8 ACL22559	ACL22559 DNA clone
43	191	40.3	693	8 ACL22560	ACL22560 DNA clone
44	191	40.3	712	8 ACL22570	ACL22570 DNA clone
45	190.4	40.2	668	3 AAC34045	Aac34045 Arabidops

## ALIGNMENTS

## RESULT 1

AAC50492

ID AAC50492 standard; DNA; 474 BP.

XX

AC AAC50492;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 65043.

XX

KW Hybridisation assay; Genetic mapping; Gene expression control;

KW protein identification; signal transduction pathway; metabolic pathway;

KW Promoter; termination sequence; 5S.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR

PR 05-MAR-1999; 99US-0123180P.

PR

PR 09-MAR-1999; 99US-0123548P.

PR

PR 23-MAR-1999; 99US-0125788P.

PR

PR 29-MAR-1999; 99US-0126264P.

PR

PR 25-MAR-1999; 99US-0126785P.

PR

PR 01-APR-1999; 99US-0127462P.

PR

PR 06-APR-1999; 99US-0128234P.

PR

PR 08-APR-1999; 99US-0128714P.

PR

PR 16-APR-1999; 99US-0129845P.

PR

PR 19-APR-1999; 99US-0130077P.

PR

PR 21-APR-1999; 99US-0130449P.

PR

PR 23-APR-1999; 99US-0130510P.

PR

PR 28-APR-1999; 99US-0130891P.

PR

PR 30-APR-1999; 99US-0131449P.

PR

PR 30-APR-1999; 99US-0132048P.

PR

PR 04-MAY-1999; 99US-0132407P.

PR

PR 05-MAY-1999; 99US-0132484P.

PR

PR 06-MAY-1999; 99US-0132485P.

PR

PR 06-MAY-1999; 99US-0132486P.

PR

PR 07-MAY-1999; 99US-0132487P.

PR

PR 07-MAY-1999; 99US-0132863P.

PR

PR 11-MAY-1999; 99US-0134256P.

PR

PR 14-MAY-1999; 99US-0134218P.

PR

PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139452P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140931P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148119P.  
PR 13-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 17-AUG-1999; 99US-0149368P.  
PR 18-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 16-SEP-1999; 99US-0154018P.  
PR 20-SEP-1999; 99US-0154039P.  
PR 22-SEP-1999; 99US-0154779P.  
PR 23-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 28-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.

```
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 474; DB 3; Length 474;
Best Local Similarity 100.0%; Pred. No. 5.7e-141;
Matches 474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCTCTAGCAGCAAACTGCAAGAAATACACGAGGAAAGCTCCGAGACTCTG 60
Db 1 ATGGCTCTAGCAGCAAACTGCAAGAAATACACGAGGAAAGCTCCGAGACTCTG 60
Qy 61 CTCGCTACCAAGGGCGGAGGAAATCTGCGCGACTCTGGAGGAGTCAAGAACTTCAC 120
Db 61 CTCGCTACCAAGGGCGGAGGAAATCTGCGCGACTCTGGAGGAGTCAAGAACTTCAC 120
Qy 121 CGTTACCGTCCCGGAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGCACAGAG 180
Db 121 CGTTACCGTCCCGGAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGCACAGAG 180
Qy 181 TTGTTGATCCGTAACCTCTTTTCAAGCTCTGTTGTTGTAATCGCTCAAGATTACAG 240
Db 181 TTGTTGATCCGTAACCTCTTTTCAAGCTCTGTTGTTGTAATCGCTCAAGATTACAG 240
Qy 241 ACGGATCTGAGATTCAGAGCCATCGGTGTTAGCTCTTCAAGAAGCTGCTGAACATAT 300
Db 241 ACGGATCTGAGATTCAGAGCCATCGGTGTTAGCTCTTCAAGAAGCTGCTGAACATAT 300
Qy 301 TTGACGGATCTGAGATTCAGAGCCATCGGTGTTAGCTCTTCAAGAAGCTGCGGAGGCA 360
Db 301 TTGACGGATCTGAGATTCAGAGCCATCGGTGTTAGCTCTTCAAGAAGCTGCGGAGGCA 360
Qy 361 TATTTGGTGGTCTTTTGAAGACAAATCTGTGTCGATTCATCAAGAGGGTTTACG 420
Db 361 TATTTGGTGGTCTTTTGAAGACAAATCTGTGTCGATTCATCAAGAGGGTTTACG 420
Qy 421 ATAATGCTTAAGATGTTCAATTTGCAAGAGGATTCGTTGAGAGCGTCTTAG 474
Db 421 ATAATGCTTAAGATGTTCAATTTGCAAGAGGATTCGTTGAGAGCGTCTTAG 474
```

```
RESULT 2
AAC50504
ID AAC50504 standard; DNA; 507 BP.
AC AAC50504;
DX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 65088.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
```

```
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
```







```
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151085P.
PR 27-AUG-1999; 99US-0151086P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 67.2%; Score 318.6; DB 3; Length 531;
Best Local Similarity 85.4%; Pred. No. 2.8e-91;
Matches 405; Conservative 0; Mismatches 5; Indels 64; Gaps 2;

QY 1 ATGGCTCTGACGAAGCAAACTGCAGAGAAATCACACGAGGAAAGCTCCGAGGACTCTG 60
DB 89 ATGGCTGTACAGAGCAAACTGCAGAGAAATCACACGAGGAAAGCTCCGAGGACTCTG 148

QY 61 CTCGCTACCAAGGCGGAGGAAATCTGCGCGACTACTGGAGGAGTCAAGAAACCTCAC 120
DB 149 CTCGCTACCAAGGCGGAGGAAATCTGCGCGACTACTGGAGGAGTCAAGAAACCTCAC 208

QY 121 CGTTACCGTCCCGAACCGTCTGCTGAGTTGTTGTAATACCAAGAGCACAGAG 180
DB 209 CGTTACCGTCCCGAACCGTCTGCTGAGTTGTTGTAATACCAAGAGCACAGAG 268

QY 181 TTGTTGATCCGTAACCTTCCTTTTCAACGCTCTGTTGTTGTAATCGCTCAAGATTACAAG 240
DB 269 TTGTTGATCCGTAACCTTCCTTTTCAACGCTCTGTTGTTGTAATCGCTCAAGATTACAAG 328
```

```
QY 241 ACGATCTGAGATTCAGAGCCATGCGTCTTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
DB 329 ACGATCTGAGATTCAGAGCCATGCGTCTTAGCTCTTCAAGAGCTGCTGAAGCATAT 382
QY 301 TTGACGATCTGAGATTCAGAGCCATGCGTCTTAGCTCTTCAAGAGCTGCGGAGGCA 360
DB 383 -----GCA 385
QY 361 TATTGGTGGTTTGTGTAAGACACAAATCTGTGCCANTCATGCAAGAGGGTTACG 420
DB 386 TATTGGTGGTTTGTGTAAGATACAAATCTGTGCCANTCATGCAAGAGGGTTACG 445
QY 421 ATATGCTTAAAGATGTTCAATTTGCGCAAGAGGATTCGTGGAGGCGTCTTAG 474
DB 446 ATCATGCTTAAAGATGTTCAATTTGCGC-AGAAGGATTCGTGGAGGCGTCTTAG 498

RESULT 4
AAC39654
ID AAC39654 standard; DNA; 621 BP.
XX AAC39654;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 25421.
XX Arabidopsis thaliana.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.
XX Arabidopsis thaliana.
PN EF1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 08-APR-1999; 99US-0128234P.
XX 16-APR-1999; 99US-0128714P.
XX 19-APR-1999; 99US-0129845P.
XX 21-APR-1999; 99US-0130077P.
XX 23-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 23-APR-1999; 99US-0130891P.
XX 28-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 06-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132863P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 14-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134768P.
XX 19-MAY-1999; 99US-0134941P.
XX 20-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135353P.
XX 24-MAY-1999; 99US-0135629P.
XX 25-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
```

PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139811P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143543P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144083P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0136782P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-015659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

```
Query Match          63.5%; Score 301.2; DB 3; Length 621;
Best Local Similarity 81.9%; Pred. No. 1.1e-85;
Matches 388; Conservative 0; Mismatches 23; Indels 63; Gaps 1;

QY 1 ATGGCTCTGACGAAGCAAACTGCAAGAAAATACACGGAGGAAAGCTCCGAGGACTCTG 60
Db 92 ATGGCTCTGACGAAGCAAACTGCAAGAAAATACACGGAGGAAAGCTCCGAGGAGCAG 151
QY 61 CTCGCTACCAAGGGCGGAGGAAATCTGGCGGACTACTGAGGAGTCAAGAAACCTCAC 120
Db 152 CTCGCTACCAAGGGCGGAGGAAATCTGCAAGGACTACTGAGGAGTCAAGAAACCTCAC 211
QY 121 CGTTACCGTCCCGGAACCGTCTTCGTGAGATTCTGTAATACCAAGAGCACAGAG 180
Db 212 CGTTTCCTCTCGAAGCGTCTGCTCTTCGTGAGATTCTGTAATACCAAGAGCACAGAG 271
QY 181 TTGTTGATCGGTAACTTCTTTTCAACGCTCTGTTGCTGAAATGCTTCAAGATTACAAG 240
Db 272 TTGTTGAAACCGTAACTTCTTCCATTCACAGCTCTGTTGCTGAAATGCTTCAAGATTCAA- 330
QY 241 ACGGATCTGAGATCCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
Db 331 ----- 330
QY 301 TTGACGGATCTGAGATCCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTGCGGAGGCA 360
Db 331 --GACGGATCTGAGATCCAAAGCCATCGAGTGTAGCTCTTCAAGAGCTGCGGAGGCA 388
QY 361 TATTGCTGGGTTGTTTGAAGACACAAATCTGTGTCATTCATGCAAGAGGTTACG 420
Db 389 TATTGCTGGGTTGTTTGAAGACACAAATCTGTGTCATTCATGCAAGAGGTTACG 448
QY 421 ATAATGCTTAAAGATGTTCAATTTGCAAGAGGATTCGTGAGAGCGTGTAG 474
Db 449 ATATGCTTAAAGATGTTCAATTTGCAAGAGGATTCGTGAGAGCGTGTAG 502

RESULT 5
AAC38179
ID AAC38179 standard; DNA; 459 BP.
AC AAC38179;
XX
XX
DT 17-OCT-2000 (first entry)
DE
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 20053.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
```

```
PR 23-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134376P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140921P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
```



PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127452P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-01334256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134376P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136382P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139432P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140891P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 09-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148119P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.



```

PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0158294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      44.0%; Score 208.6; DB 3; Length 411;
Best Local Similarity 79.4%; Pred. No. 3.6e-56;
Matches 247; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1 ATGGCTCTAGCAGCAAACTGGCAAGAAATACACGGAGGAAAGCTCCGAGGACTCTG 60
Db 1 ATGGCTCTAGCAGCAAACTGGCAAGAAATACACGGAGGAAAGCTCCGAGGACTCTG 60
QY 61 CTCGCTACCAAGGGCGGAGGAAATCTGGCGGACTACTGAGGAGTCAAGAACTTAC 120
Db 61 CTCGCTACCAAGGGCGGAGGAAATCTGGCGGACTACTGAGGAGTCAAGAACTTAC 120
QY 121 CGTTACCGTCCCGGAACCGTCTGCTGAGATTCGTAATACCAAGAGCACAG 180
Db 121 CGTTACCGTCCCGGAACCGTCTGCTGAGATTCGTAATACCAAGAGCACAG 180
QY 181 TTGTTGATCCGTAACCTTCTTCAACGCTCTTGTCTGTAATCGCTCAAGATTACA 240
Db 181 TTGTTGATCCGTAACCTTCTTCAACGCTCTTGTCTGTAATCGCTCAAGATTACA 240
QY 241 ACGGATCTGAGATTCAGAGCATCGCTGTTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
Db 241 ACGGATCTGAGATTCAGAGCATCGCTGTTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
QY 301 TTGACGGATCT 311
Db 301 TTGACGGATCT 311

```

## RESULT 7

AAC51199  
ID AAC51199 standard; DNA; 643 BP.

AC AAC51199;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 67632.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway; metabolic pathway;

KW promoter; termination sequence; ss.

XX Arabidopsis thaliana.  
OS EF1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-00301439.  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.

PR	29-JUN-1999;	99US-0140981P.	PR	13-SEP-1999;	99US-0153758P.
PR	30-JUN-1999;	99US-0141287P.	PR	15-SEP-1999;	99US-0154018P.
PR	01-JUL-1999;	99US-0141842P.	PR	16-SEP-1999;	99US-0154039P.
PR	01-JUL-1999;	99US-0142154P.	PR	20-SEP-1999;	99US-0154779P.
PR	02-JUL-1999;	99US-0142055P.	PR	22-SEP-1999;	99US-0155139P.
PR	06-JUL-1999;	99US-0142390P.	PR	23-SEP-1999;	99US-0155486P.
PR	08-JUL-1999;	99US-0142803P.	PR	24-SEP-1999;	99US-0155659P.
PR	09-JUL-1999;	99US-0142920P.	PR	28-SEP-1999;	99US-0156458P.
PR	12-JUL-1999;	99US-0142977P.	PR	29-SEP-1999;	99US-0156596P.
PR	13-JUL-1999;	99US-0143542P.	PR	04-OCT-1999;	99US-0157117P.
PR	14-JUL-1999;	99US-0143624P.	PR	05-OCT-1999;	99US-0157753P.
PR	15-JUL-1999;	99US-0144005P.	PR	06-OCT-1999;	99US-0157865P.
PR	16-JUL-1999;	99US-0144085P.	PR	07-OCT-1999;	99US-0158029P.
PR	16-JUL-1999;	99US-0144086P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144326P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144331P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144334P.	PR	14-OCT-1999;	99US-0159329P.
PR	20-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144332P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144632P.	PR	14-OCT-1999;	99US-0159637P.
PR	20-JUL-1999;	99US-0144884P.	PR	14-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0144814P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0145086P.	PR	21-OCT-1999;	99US-0160741P.
PR	22-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160767P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145132P.	PR	21-OCT-1999;	99US-0160814P.
PR	22-JUL-1999;	99US-0145145P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160980P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160981P.
PR	26-JUL-1999;	99US-0145276P.	PR	22-OCT-1999;	99US-0160989P.
PR	27-JUL-1999;	99US-0145913P.	PR	25-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161405P.
PR	27-JUL-1999;	99US-0145919P.	PR	25-OCT-1999;	99US-0161406P.
PR	28-JUL-1999;	99US-0145951P.	PR	26-OCT-1999;	99US-0161359P.
PR	02-AUG-1999;	99US-0146386P.	PR	26-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0146388P.	PR	26-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0146389P.	PR	28-OCT-1999;	99US-0161920P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161992P.
PR	04-AUG-1999;	99US-0147204P.	PR	28-OCT-1999;	99US-0161993P.
PR	05-AUG-1999;	99US-0147302P.	PR	29-OCT-1999;	99US-0162142P.
PR	05-AUG-1999;	99US-0147192P.			
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			
PR	08-AUG-1999;	99US-0147433P.			
PR	09-AUG-1999;	99US-0147935P.			
PR	10-AUG-1999;	99US-0148171P.			
PR	11-AUG-1999;	99US-0148319P.			
PR	12-AUG-1999;	99US-0148341P.			
PR	13-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-0149368P.			
PR	17-AUG-1999;	99US-0149175P.			
PR	18-AUG-1999;	99US-0149426P.			
PR	20-AUG-1999;	99US-0149722P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149929P.			
PR	23-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	25-AUG-1999;	99US-0150566P.			
PR	26-AUG-1999;	99US-0150884P.			
PR	27-AUG-1999;	99US-0151085P.			
PR	27-AUG-1999;	99US-0151066P.			
PR	30-AUG-1999;	99US-0151080P.			
PR	30-AUG-1999;	99US-0151303P.			
PR	01-SEP-1999;	99US-0151438P.			
PR	01-SEP-1999;	99US-0151930P.			
PR	07-SEP-1999;	99US-0152363P.			
PR	10-SEP-1999;	99US-0153070P.			
			Query Match 44.0%; Score 208.6; DB 3; Length 643;		
			Best Local Similarity 79.4%; Pred. No. 4.4e-56;		
			Matches 247; Conservative 0; Mismatches 64; Indels 0; Gaps 0;		
QY	1	ATGGCTCGTACGAAGCAAACTGCAAGAAATCACCGAGGAAAGCTCCGAGGACTCTG 60			
DB	91	ATGGCTCGTACGAAGCAAACTGCAAGAAATCACCGAGGAAAGCTCCGAGGACTCTG 150			
QY	61	CTCGTACCAAGCGCGGAGGAAATCTGCGCGGACTACTGGAGGAGTCAAGAACTCAC 120			
DB	151	CTTGCTACTAAGGCTCTGCTAAATCTGCACCAACTACTTGGTGGACTCAAGAACTCAT 210			
QY	121	CGTTACCGTCCGAGACCGTCTTCTGAGATTCGTAATACCAAGAGCAGCAGAG 180			
DB	211	CGTTACCGTCTGGAATCTGCTCTCCGTGAATCCGTAAGTACCAAGAGGAGTACTGAG 270			
QY	181	TTGTTGATCCGTAATCTTCTTTTCAACGCTCTTGTGTAATCGCTCAAGATTACAAG 240			
DB	271	TTGCTTATCAGGAATCTGCAATTCAGAGGCTAGTCCGTTGAGATTCCTCAAGATTCAAG 330			
QY	241	ACGATCTGAGATTCAGAGCAGCAGGCTGTTAGCTCTTCAAGAGCTGCTGAAGCATAT 300			
DB	331	ACTGATTGCTTCCAGAGCAGCAGGCTGCTTCTAGCTCTCCAGAGCTGCTGAGAGCATAT 390			
QY	301	TTGACCGATCT 311			
DB	391	CTTGTTGGTCT 401			



```
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149902P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151068P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159337P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 44.0%; Score 208.6; DB 3; Length 674;
Best Local Similarity 79.4%; Pred. No. 4.5e-56;
Matches 247; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACCGAGGAGAAAGCTCCGAGGACTCTG 60
DB 85 ATGGCTCGTACGAAGCAAACTGCTAAGTCCACCGAGGTAAGCTCAAGGAGCA 144
QY 61 CTCGCTACCAAGCGCGGAGGAATCTGCCCGACTACTGGAGGAGTCAAGAAACCTCAC 120
DB 145 CTCGCTACTAAGCTCTCGTAAATCTGCACCAACTACTGCTGGAGTCAAGAAACCCAT 204
QY 121 CGTTACCGTCCCGAACCGCTCGCTCTTCGTGAGATTCGTAATAACCAAGAGACAGAG 180
DB 205 CGTTACCGTCCCGAACCGTCTCTCCGTGAAATCCGTAAGTACCAGAGAGTACTGAG 264
```

```
QY 181 TTGTGATCCGTAACCTTCCTTTTCAACGCTTGTTCGTGAATCGCTCAAGATTCAAG 240
DB 265 TTGCTTATCAGGAACCTGATTTTCAGAGGCTAGTCGCTGAGATTGCTCAAGATTCAAG 324
QY 241 ACGGATCTGAGATTCAGAGCCATGCGGTGTTAGCTCTTCAAGAAAGCTGCTGAAGCATAT 300
DB 325 ACTGATTGCGTTCCAGAGCCATGCTGTCTTAGCTCTCCAGGAAGCTGCAAGAGCATAT 384
QY 301 TTGACGGATCT 311
DB 385 CTGTGTTGCTCT 395

RESULT 9
AAC32694
ID AAC32694 standard; DNA; 677 BP.
XX
AC AAC32694;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 305.
XX
KW Hybridisation assay; Genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
PF 25-FEB-2000; 2000EP-00301439.
PR 25-FEB-1999; 99US-0121825P.
PR 09-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
```

PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140931P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143242P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 18-JUL-1999; 99US-0144325P.  
PR 18-JUL-1999; 99US-0144331P.  
PR 18-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144332P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145132P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 25-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.

PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148585P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155565P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 44.0%; Score 208.6; DB 3; Length 677;  
Best Local Similarity 79.4%; Pred. No. 4.5e-56;  
Matches 247; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
QY      1  ATGGCTCTGACGAGCAAACTGCAAGAAAAATCACACGAGGAGAAAGCTCCGAGGACTCTG  60
Db      86  ATGGCTCTGACGAGCAAACTGCAAGAAAAATCACACGAGGAGTAAAGCTCCGAGGAGCA  145
QY      61  CTCGCTACCAAGGCGGAGGAAATCTGCGCGGACTCTGCGGAGGAGTCAAGAAACCTTCA  120
Db     146  CTTGCTACTAGGCTGCTCGTAAATCTGCACCACTACTGGTGGAGTCAAGAAACCAT  205
QY     121  CGTTACCGTCCGGAACCGTGGCTCTTCGTGAGATTCTGTAATACCAAGAGCAGACAG  180
Db     206  CGTTACCGTCCGGAACCTGCTCTCGTGAATCCGTAAGTACCAAGAGTACTGAG  265
QY     181  TTGTTGATCCGTAACCTCTCTTTCAAGCTCTGTTGCTGTAATCGCTCAAGATTACAAG  240
Db     266  TTGCTTATCAGAAACTCCATTTTCAGAGGCTAGTCCGTTGAGATTGCTCAAGATTCAAG  325
QY     241  ACGGATCTGAGATTCCAGAGCCATCGGCTGTTAGCTCTTCAAGAGCTGCTGAAGCATAT  300
Db     326  ACTGATTGCGTTTCCAGAGCCATCTGTTAGCTCTTCCAGGAGCTGCAGAGCATAT  385
QY     301  TTGACGGATCT 311
Db     386  CTTGTTGGTCT 396

RESULT 10
AAC47929
ID    AAC47929 standard; DNA; 1005 BP.
AC
XX
AC    AAC47929;
XX
DT    18-OCT-2000 (first entry)
XX
XX
DE    Arabidopsis thaliana DNA fragment SEQ ID NO: 55633.
XX
XX
KW    Hybridisation assay; Genetic mapping; gene expression control;
KW    protein identification; signal transduction pathway; metabolic pathway;
KW    promoter; termination sequence; ss.
XX
OS    Arabidopsis thaliana.
XX
FN    EP1033405-A2.
XX
PD    06-SEP-2000.
XX
PF    25-FEB-2000; 2000EP-00301439.
XX
XX
25-FEB-1999; 99US-0121825P.
PR
05-MAR-1999; 99US-0123180P.
PR
09-MAR-1999; 99US-0123548P.
PR
23-MAR-1999; 99US-0125788P.
PR
25-MAR-1999; 99US-0126264P.
PR
29-MAR-1999; 99US-0126785P.
PR
01-APR-1999; 99US-0127462P.
PR
06-APR-1999; 99US-0128234P.
PR
08-APR-1999; 99US-0128714P.
PR
16-APR-1999; 99US-0129845P.
PR
19-APR-1999; 99US-0130077P.
PR
21-APR-1999; 99US-0130449P.
PR
23-APR-1999; 99US-0130510P.
PR
28-APR-1999; 99US-0130891P.
PR
30-APR-1999; 99US-0131449P.
PR
30-APR-1999; 99US-0132048P.
PR
04-MAY-1999; 99US-0132484P.
PR
05-MAY-1999; 99US-0132485P.
PR
06-MAY-1999; 99US-0132486P.
PR
06-MAY-1999; 99US-0132487P.
PR
07-MAY-1999; 99US-0132863P.
PR
11-MAY-1999; 99US-0132863P.
PR
14-MAY-1999; 99US-0134218P.
PR
14-MAY-1999; 99US-0134219P.
PR
14-MAY-1999; 99US-0134221P.
PR
14-MAY-1999; 99US-0134370P.
PR
18-MAY-1999; 99US-0134768P.
PR
19-MAY-1999; 99US-0134941P.
PR
20-MAY-1999; 99US-0135124P.
PR
21-MAY-1999; 99US-0135353P.
PR
21-MAY-1999; 99US-0135629P.
PR
25-MAY-1999; 99US-0136021P.
PR
27-MAY-1999; 99US-0136782P.
PR
28-MAY-1999; 99US-0137222P.
PR
01-JUN-1999; 99US-0137528P.
PR
03-JUN-1999; 99US-0137528P.
PR
04-JUN-1999; 99US-0137724P.
PR
07-JUN-1999; 99US-0138094P.
PR
10-JUN-1999; 99US-0138540P.
PR
10-JUN-1999; 99US-0138847P.
PR
14-JUN-1999; 99US-0139119P.
PR
16-JUN-1999; 99US-0139452P.
PR
17-JUN-1999; 99US-0139452P.
PR
18-JUN-1999; 99US-0139454P.
PR
18-JUN-1999; 99US-0139455P.
PR
18-JUN-1999; 99US-0139456P.
PR
18-JUN-1999; 99US-0139457P.
PR
18-JUN-1999; 99US-0139458P.
PR
18-JUN-1999; 99US-0139459P.
PR
18-JUN-1999; 99US-0139460P.
PR
18-JUN-1999; 99US-0139461P.
PR
18-JUN-1999; 99US-0139462P.
PR
18-JUN-1999; 99US-0139463P.
PR
18-JUN-1999; 99US-0139750P.
PR
18-JUN-1999; 99US-0139763P.
PR
21-JUN-1999; 99US-0139817P.
PR
22-JUN-1999; 99US-0139899P.
PR
23-JUN-1999; 99US-0140353P.
PR
23-JUN-1999; 99US-0140354P.
PR
24-JUN-1999; 99US-0140695P.
PR
28-JUN-1999; 99US-0140823P.
PR
29-JUN-1999; 99US-0140991P.
PR
30-JUN-1999; 99US-0141287P.
PR
01-JUL-1999; 99US-0141842P.
PR
01-JUL-1999; 99US-0142154P.
PR
02-JUL-1999; 99US-0142055P.
PR
06-JUL-1999; 99US-0142390P.
PR
08-JUL-1999; 99US-0142803P.
PR
09-JUL-1999; 99US-0142920P.
PR
12-JUL-1999; 99US-0142977P.
PR
13-JUL-1999; 99US-0143542P.
PR
14-JUL-1999; 99US-0143624P.
PR
15-JUL-1999; 99US-0144005P.
PR
16-JUL-1999; 99US-0144085P.
PR
16-JUL-1999; 99US-0144086P.
PR
19-JUL-1999; 99US-0144325P.
PR
19-JUL-1999; 99US-0144331P.
PR
19-JUL-1999; 99US-0144332P.
PR
19-JUL-1999; 99US-0144333P.
PR
19-JUL-1999; 99US-0144334P.
PR
19-JUL-1999; 99US-0144335P.
PR
20-JUL-1999; 99US-0144352P.
PR
20-JUL-1999; 99US-0144632P.
PR
20-JUL-1999; 99US-0144884P.
PR
21-JUL-1999; 99US-0144814P.
PR
21-JUL-1999; 99US-0145086P.
PR
21-JUL-1999; 99US-0145088P.
PR
22-JUL-1999; 99US-0145085P.
PR
22-JUL-1999; 99US-0145087P.
PR
22-JUL-1999; 99US-0145089P.
PR
22-JUL-1999; 99US-0145192P.
PR
23-JUL-1999; 99US-0145145P.
PR
23-JUL-1999; 99US-0145218P.
PR
23-JUL-1999; 99US-0145224P.
PR
26-JUL-1999; 99US-0145276P.
```

```
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149829P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.

PR 25-OCT-1999; 99US-0161405P.
PR 26-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 43.8%; Score 207.6; DB 3; Length 1005;
Best Local Similarity 79.1%; Pred. No. 1.1e-55;
Matches 246; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 ATGGCTCTGTAAGCAAACTGCAAGAAAATCACAGGAGGAAAGCTCGAGGACTCTG 60
DB 373 ATGGCTCTGTAAGCAAACTGTAAGTCCACCGGAGGTAAGCTCAAGGAAGCAA 432
QY 61 CTCGCTACCAAGCGCGGAGGAAATCTGCGCGACTACTGGAGGAGTCAAGAACTCAC 120
DB 433 CTGTCTACTAAGCTGCTGTAATACTGCACCAACNACTGTGGAGTCAAGAACCAAT 492
QY 121 CGTTACCGTCCCGGAACCGTCTCTTCTGTGAGATTCTGTAATAACAGAGAGCAAGAG 180
DB 493 CGTTACCGTCCCGGAACCTGTTCTCTCGTGAATCCGTAAGTACCAGAGAGTACTGAG 552
QY 181 TTGTTGATCCGTAATACTCTCTTTTCAACGCTCTTGTTCGTGAAATCGCTCAAGATTACAAG 240
DB 553 TTGCTTATCAGGAAATGCCAATTCAGAGGCTAGTCCGTGAGATTGCTCAAGATTCAAG 612
QY 241 ACGGATCTGAGATTCCAGAGCCATCGGCTGTTAGCTCTTCAAGAGCTCTCGAAGCATAT 300
DB 613 ACTGATTGCGTTTCAGAGCCATGCTGTCTTAGCTCTCCAGAGAGCTCGAAGCATAT 672
QY 301 TTGACGGATCT 311
DB 673 CTGTGTGTCT 683

RESULT 11
AAC33712
ID AAC33712 standard; DNA; 646 BP.
XX
AC AAC33712;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 4049.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-012180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P;
```





```

PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 43.7%; Score 207; DB 3; Length 646;
Best Local Similarity 79.1%; Pred. No. 1.4e-55;
Matches 246; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 ATGGCTCGTACGAGCAAACTGCAAGAAATACACGGAGGAAAGCTCCGAGGACTCTG 60
DB 94 ATGGCTCGTACGAGCAAACTGCAAGAAATACACGGAGGAAAGCTCCGAGGAGCAA 153

QY 61 CTCGCTACCAAGCGCGAGGAATCTGCCGACTACTCGAGGACTCAAGAACTCTAC 120
DB 154 CTTGCTACTAAGGCTCTCGTAAATCTGCACCAACTACTGGTGGTGAAGCAACCAT 213

QY 121 CGTTACCGTCCGGAACCGTGGCTCTTCGTGAGATTGCTAAATACCAAGAGCAGAG 180
DB 214 CGTTACCGTCTGGAATCTGCTCCGTGAATCGTAAATACCAAGAGACTCTGAG 273

QY 181 TTGTTGATCGTAAATCTCTTTTCAAGCTCTGTTGCTGGAATCGTCAAGATTACAAG 240
DB 274 TTGCTTATCAGGAAGCTGCCATTTTCAGAGCTAGTCCGTTGAGATTGCTCAAGATTCAAG 333

QY 241 ACGGATCTGAGATTCCAGAGCCATCGGTTGAGTCTTCAAGAGCTGCTGAAGCATAT 300
DB 334 ACTGATTGGCTTCAGAGCCATGCTGCTAGCTCTCCAGGAGCTGCAGAGCATAT 393

QY 301 TTGACGGATCT 311
DB 394 CTTGTTGGTCT 404

RESULT 12
ADE81685
ID ADE81685 standard; cDNA; 499 BP.
AC ADE81685;
XX
XX
XX 29-JAN-2004 (first entry)
DE Arabidopsis thaliana expressed polynucleotide seq id 456.
KW genetically modified organism; transgenic organism; plant;
KW inhibitor testing; activator testing; modifier testing; fungicide;
KW insecticide; genetic function; genetic regulation; cellular metabolism;
XX gene; ss.
XX Arabidopsis thaliana.
XX
XX US2003115639-A1.
XX
XX 19-JUN-2003.
XX

```

```

PF 26-JAN-2001; 2001US-00770961.
XX
PR 27-JAN-2000; 2000US-0178466P.
XX
XX
PA (GORL/) GORLACH J.
PA (ANY/) AN Y.
PA (HAMI/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (RAIN/) RAINES T W.
PA (YUY/) YU Y.
PA (RAME/) RAVEAKA J G.
PA (PAGE/) PAGE A.
PA (MATH/) MATHW A V.
PA (LEDF/) LEDFORD B L.
PA (WOES/) WOESSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
PA (KRIC/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K R.
PA (ALLE/) ALLEN K.
PA (HOFF/) HOFFMAN N.
PA (HURB/) HURBAN P.
XX
XX Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
PI Hurban P;
XX
XX WPI; 2003-810930/76.
XX
PT Novel Arabidopsis thaliana nucleic acids useful for generating
PT genetically modified transgenic organisms, for screening biologically
PT active agents such as fungicides, insecticides.
XX
XX Claim 1; SEQ ID NO 456; 44bp; English.
XX
XX The invention describes a nucleic acid (I) comprising a sequence capable
XX of hybridising under stringent conditions to any one of 99 fully defined
XX Arabidopsis thaliana sequences (I) as given in specification e.g., 360,
XX 1137, 455, 219, 472 nucleotides, etc, or its fragment. (I) is useful as a
XX hybridisation probe to complementary molecules in a cDNA library. (I) is
XX also useful for generating genetically modified and transgenic organisms,
XX usually plant cells and plants. A protein encoded by (I) is useful in
XX screening assays to determine the effect of candidate inhibitors,
XX activators or modifiers of the gene product. The protein is also useful
XX for screening biologically active agents e.g., fungicides and
XX insecticides. A genetically modified cell, comprising an exogenous
XX nucleic acid, where the nucleic acid comprises transcription regulatory
XX sequences operably linked to a sequence capable of hybridising under
XX stringent conditions to (I) is useful in the study of genetic function
XX and regulation, for alteration of the cellular metabolism and for
XX screening compounds that may affect the biological function of the gene
XX or gene product. This sequence represents an Arabidopsis thaliana
XX polynucleotide of the invention.
XX
XX Sequence 499 BP; 146 A; 121 C; 111 G; 116 T; 0 U; 5 Other;
XX
XX Query Match 43.3%; Score 205.2; DB 9; Length 499;
XX Best Local Similarity 78.4%; Pred. No. 4.8e-55;
XX Matches 243; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
XX
QY 2 TGCGCTCGTACGAGCAAACTGCAAGAAATACACGGAGGAAAGCTCCGAGGACTCTGC 61
DB 80 TGNNTGNCCAGCAAAACCGCTCGTAACTCCACCGAGGTAAGCTCCAGGAAGCAAC 139
QY 62 TCCTACCAAGCGCGAGGAAATCTGCCGCTACTCGAGGAGTCAAGAAACCTCAAC 121
DB 140 TTGCTACTAAGGCTGCTCGTAAATCTGCACCAACTACTGCTGGAGTCAAGAAACCATC 199
QY 122 GTTACCGTCCGGAACCGTGGCTCTTCGTGAGATTGTTAAATACCAAGAGCAGAGT 181
DB 200 GTTACCGTCCGGAACCTGTTGCTCTCCGTGAATCCGTAAGTACCAGAGAGTACTGAGT 259

```

QY 182 TGTGATCCGTAACCTTCCTTTTCAACGCTTGTTCGTGAATCGCTCAAGATTACAGA 241  
 DB |||||  
 DB 260 TGCATTATCAGGAACGCAATTCAGAGGCTAGTTTCGTGAGATTCTCAAGATTCAAGA 319  
 QY 242 CGGATCTCAGATTCCAGAGCCATCGGTGTTAGCTCTTCAAGAAGCTGCTGAAGCATATT 301  
 DB |||||  
 DB 320 CTGATTTCGGTTTCAGAGCCATGCTGCTTACCTCTCCAGGAAGCTGCAGNAGCATATC 379  
 QY 302 TGAAGGATCT 311  
 DB |||||  
 DB 380 TTGTTGGTCT 389

RESULT 13  
 ID ABK31041  
 AC ABK31041 standard; cDNA; 699 BP.  
 XX  
 XX 07-AUG-2003 (revised)  
 DT 23-APR-2002 (first entry)  
 XX  
 XX Plant dwarfing/stunting related cDNA seq ID 455.  
 XX  
 KW Plant; ss; dwarfism; stunting; EST; expressed sequence tag;  
 KW transgenic plant; plant metabolism.  
 OS Unidentified.  
 XX  
 PN WO200208410-A2.  
 XX  
 PD 31-JAN-2002.  
 XX  
 PF 20-JUL-2001; 2001WO-US023120.  
 XX  
 PR 20-JUL-2000; 2000US-0219809P.  
 PR 20-JUL-2000; 2000US-0219810P.  
 XX  
 XX (DOWC ) DOW CHEM CO.  
 PA (REDDY) REDDY S. A.  
 PA (LARR) LARRINDA M I.  
 PA (RUEG) RUEGGER M.  
 PA (WEG) WEGIERZ T.  
 PA (BLAK) BLAKESLEE B.  
 PA (ORIE) ORLEDO V B J.  
 PA (SAVI) SAVICKAS J P.  
 PA (MCCR) MCCREY A D.  
 PA (MILL) MILLER A B.  
 PA (POGU) POGUE P G.  
 PA (DELL) DELLA-CIOPPA R G.  
 PA (WOLF) WOLFE M G.  
 PA (ZHEN) ZHENG W.  
 PA (GACH) GACHOTTE D.  
 PA (GROS) GROSLEY R.  
 PA (PELL) PELL R.

Reddy SA, Larrinua MI, Ruegger M, Weglarz T, Blakeslee B;  
 Orleodo VBJ, Savickas JP, McCreery AD, Miller AB, Pogue PG;  
 Della-Cioppa RG, Wolfe MG, Zheng W, Gachotte D, Grosley R, Pell R;  
 WPI; 2002-164823/21.

Polynucleotide and amino acid sequences identified in one or more  
 PT metabolic pathways that lead to dwarfism and stunting in plants, useful  
 PT in agriculture to create dwarf varieties of any plant species.  
 XX  
 PS Disclosure; Fig 9; 717pp; English.  
 XX  
 CC The invention relates to polynucleotide and amino acid sequences  
 CC identified in one or more metabolic pathways that lead to dwarfism and  
 CC stunting in plants. Also included are vectors comprising the  
 CC polynucleotides, transgenic plants (including the seed and leaf)

CC transfected with the polynucleotides or vectors, a process for altering  
 CC the metabolism of a plant comprising providing the above vector and a  
 CC plant, and transfecting the plant with the vector under conditions such  
 CC that the metabolism of the plant is altered by expression of the isolated  
 CC nucleic acid from the vector, e.g. such that a stunting phenotype in an  
 CC industrial plant is produced and a process for the characterisation of  
 CC fractionated biological samples, comprising (a) providing one or more  
 CC fractionated biological samples, references samples, a gas chromatography  
 CC apparatus, a mass spectroscopy apparatus or data analysis software and  
 CC (b) treating the fractionated biological samples and the reference  
 CC samples with the gas chromatography apparatus to generate chromatographic  
 CC data corresponding to the fractionated biological samples and the  
 CC reference samples, (c) treating the fractionated biological samples and  
 CC the reference samples with the mass spectroscopy apparatus to generate  
 CC spectroscopic data corresponding to the fractionated biological samples  
 CC and the reference samples and (d) processing the chromatographic and the  
 CC spectroscopic data with the data analysis software. The nucleic acid and  
 CC the vector are useful for altering the metabolism of a plant and for  
 CC stunting a plant. The nucleic acids are useful in agriculture to create  
 CC dwarf varieties of any plant species. The present sequence is a plant  
 CC cDNA contig or singleton (related to dwarfism/stunting) identified by  
 CC searching a nucleic acid database with plant EST (expressed sequence tag)  
 CC and a BLAST (basic local alignment tool) stringency of e-20. (Updated on  
 CC 07-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 699 BP; 189 A; 130 C; 178 G; 201 T; 0 U; 1 Other;

Query Match 42.4%; Score 200.8; DB 6; Length 699;  
 Best Local Similarity 78.0%; Pred. No. 1.4e-53;  
 Matches 241; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 3 GGCTCGTACGAGCAAACTGCAAGAAATCACACGGAGGAAAGCTCCGRGACTCTGCT 62  
 DB |||||  
 DB 64 GNCCTGCTACCAAGCAAACTGCTCGTAAGTCTACCGAGGAAAGGACCTAGGAAGCAACT 123  
 QY 63 CGCTACCAAGGCGGAGGAAATCTGCGCCGACTACTGGAGGAGTCAAGAAACCTCACCG 122  
 DB |||||  
 DB 124 TGCTACTAAGGCTGCTCGTAAGTCTGCTCTCTCTACTACTGCTGAGTAAAGAACTCACAG 183  
 QY 123 TTACGTCCTCCGAACCGTCGCTCTTCGTGAGATTCGTAAATACCAAGAGACAGAGTT 182  
 DB |||||  
 DB 184 ATACCGCCCTGCTACTGTTGCTCTTCGTGAAATCCGTAAGTACCAGAGAGTACTGAGCT 243  
 QY 183 GTTGATCCGTAAACTTCCTTTTCAACGCTTGTGTTGCTCAAGATTCGCTCAAGATTACAAGAC 242  
 DB |||||  
 DB 244 CTTGATCAGGAAGCTCCCATTCAGAGGCTTGTTCGTGAATTCCTCAGGATTTCAAGAC 303  
 QY 243 GGATCTGAGATTCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTGCTGAAGCATATT 302  
 DB |||||  
 DB 304 TGATCTGGTTCAGAGATCATGCTGTTGGCTCTTCAGGAGGCTGCTGAGGCATCTT 363  
 QY 303 GACGATCT 311  
 DB |||||  
 DB 364 GGTGGTCT 372

RESULT 14  
 AAC51198  
 ID AAC51198 standard; DNA; 714 BP.  
 XX  
 XX AAC51198;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 67629.  
 XX  
 KW Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway; metabolic pathway;  
 KW promoter; termination sequence; ss.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.

XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PF 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134376P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.

PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144684P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.

```
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160815P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160982P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 42.0%; Score 199; DB 3; Length 714;
Best Local Similarity 77.5%; Pred. No. 5.4e-53;
Matches 241; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 1 ATGGCTCGTACGAAGCAAACTGCAGAAAATCACACGAGGAGAAAGCTCCGAGGACTCTG 60
   |||||
Db 64 ATGGCTCGTACTAGCAACACGCTGTAGTCTACTGGAGGAAAGGCTCCTAGGAGCAG 123
   |||||

QY 61 CTCGCTACCAAGCGGCGGAGGAAATCTGCCGACTTCTGGAGGAGTCAAGAAACCTCAC 120
   |||||
Db 124 CTTGCTACAAAGGCTGCAGTAAGTCTGCACCAACCACTGGAGGAGTCAAGAAAGCCCAT 183

QY 121 CGTTACCGTCCGGAACCGCTCTTCTGTGAGATTCGTAATACCAGAGAGCACAGAG 180
   |||||
Db 184 CGTTACCGTCCGGAACCGCTCTTGCACCTACGCTGAAATTCGTAATACCAGAGAGTACCGAG 243
   |||||

QY 181 TTGTGATCCGTAACCTCTCTTTTCAACGCTTGTGTTGTAATCGCTCAAGATTACAAG 240
   |||||
Db 244 TTGCTGATCAGGAGCTCCCTTTCCAGAGGCTAGTTCGTGAGATTGCCAGGATTTCAG 303
   |||||

QY 241 ACGGATCTGAGATTCAGAGCCATCGGTTGATGCTTTCAAGAGCTCTGTAAGCATAT 300
   |||||
Db 304 ACTGACTTGGTTTCCAGAGCCATGCTGTGCTTGCACTCCAGGAGGCTGCTGAAGCATAC 363
   |||||

QY 301 TTGACCGATCT 311
   |||||
Db 364 CTGTGGGTCT 374
   |||||

RESULT 15
AAC34378
ID AAC34378 standard; DNA; 717 BP.
XX
```

```
AC AAC34378;
XX 17-OCT-2000 (first entry)
DT Arabidopsis thaliana DNA fragment SEQ ID NO: 6434.
XX Arabidopsis thaliana.
KW Hybridisation assay; Genetic mapping; Gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN EP1033405-A2.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
```

Tue Jun 29 08:37:07 2004

PR	18-JUN-1999;	99US-01399453P;
PR	18-JUN-1999;	99US-01397850P;
PR	18-JUN-1999;	99US-0139763P;
PR	21-JUN-1999;	99US-0139817P;
PR	22-JUN-1999;	99US-01403899P;
PR	23-JUN-1999;	99US-0140353P;
PR	23-JUN-1999;	99US-0140354P;
PR	24-JUN-1999;	99US-0140695P;
PR	24-JUN-1999;	99US-0140823P;
PR	28-JUN-1999;	99US-0140991P;
PR	29-JUN-1999;	99US-0141287P;
PR	30-JUN-1999;	99US-0141842P;
PR	01-JUL-1999;	99US-0142154P;
PR	02-JUL-1999;	99US-0142055P;
PR	06-JUL-1999;	99US-0142390P;
PR	08-JUL-1999;	99US-0142803P;
PR	09-JUL-1999;	99US-0142920P;
PR	12-JUL-1999;	99US-0142977P;
PR	13-JUL-1999;	99US-0143542P;
PR	14-JUL-1999;	99US-0143624P;
PR	15-JUL-1999;	99US-0144005P;
PR	16-JUL-1999;	99US-0144086P;
PR	16-JUL-1999;	99US-0144086P;
PR	19-JUL-1999;	99US-0144325P;
PR	19-JUL-1999;	99US-0144331P;
PR	19-JUL-1999;	99US-0144332P;
PR	19-JUL-1999;	99US-0144333P;
PR	19-JUL-1999;	99US-0144334P;
PR	19-JUL-1999;	99US-0144335P;
PR	20-JUL-1999;	99US-0144352P;
PR	20-JUL-1999;	99US-0144632P;
PR	20-JUL-1999;	99US-0144884P;
PR	21-JUL-1999;	99US-0144814P;
PR	21-JUL-1999;	99US-0145086P;
PR	21-JUL-1999;	99US-0145088P;
PR	22-JUL-1999;	99US-0145085P;
PR	22-JUL-1999;	99US-0145087P;
PR	22-JUL-1999;	99US-0145089P;
PR	22-JUL-1999;	99US-0145192P;
PR	23-JUL-1999;	99US-0145145P;
PR	23-JUL-1999;	99US-0145218P;
PR	23-JUL-1999;	99US-0145224P;
PR	26-JUL-1999;	99US-0145376P;
PR	27-JUL-1999;	99US-0145913P;
PR	27-JUL-1999;	99US-0145918P;
PR	27-JUL-1999;	99US-0145919P;
PR	27-JUL-1999;	99US-0145951P;
PR	28-JUL-1999;	99US-0146386P;
PR	02-AUG-1999;	99US-0146388P;
PR	02-AUG-1999;	99US-0146389P;
PR	02-AUG-1999;	99US-0147038P;
PR	03-AUG-1999;	99US-0147204P;
PR	04-AUG-1999;	99US-0147302P;
PR	04-AUG-1999;	99US-0147392P;
PR	05-AUG-1999;	99US-0147260P;
PR	06-AUG-1999;	99US-0147303P;
PR	06-AUG-1999;	99US-0147416P;
PR	09-AUG-1999;	99US-0147493P;
PR	09-AUG-1999;	99US-0147935P;
PR	10-AUG-1999;	99US-0148319P;
PR	11-AUG-1999;	99US-0148319P;
PR	13-AUG-1999;	99US-0148341P;
PR	13-AUG-1999;	99US-0148565P;
PR	13-AUG-1999;	99US-0148684P;
PR	16-AUG-1999;	99US-0149368P;
PR	17-AUG-1999;	99US-0149175P;
PR	18-AUG-1999;	99US-0149426P;
PR	20-AUG-1999;	99US-0149722P;
PR	20-AUG-1999;	99US-0149723P;
PR	20-AUG-1999;	99US-0149923P;
PR	23-AUG-1999;	99US-0149902P;
PR	23-AUG-1999;	99US-0149930P;
PR	25-AUG-1999;	99US-0150566P;

26-AUG-1999;	99US-0150884P;
27-AUG-1999;	99US-0151056P;
27-AUG-1999;	99US-0151056P;
27-AUG-1999;	99US-0151080P;
30-AUG-1999;	99US-0151303P;
31-AUG-1999;	99US-0151438P;
01-SEP-1999;	99US-0151930P;
07-SEP-1999;	99US-0152363P;
10-SEP-1999;	99US-0153070P;
13-SEP-1999;	99US-0153758P;
16-SEP-1999;	99US-0154018P;
16-SEP-1999;	99US-0154039P;
20-SEP-1999;	99US-0154779P;
22-SEP-1999;	99US-0155139P;
23-SEP-1999;	99US-0155486P;
24-SEP-1999;	99US-0155659P;
28-SEP-1999;	99US-0155458P;
29-SEP-1999;	99US-0155696P;
04-OCT-1999;	99US-0157111P;
05-OCT-1999;	99US-0157535P;
06-OCT-1999;	99US-0157865P;
07-OCT-1999;	99US-0158029P;
08-OCT-1999;	99US-0158332P;
12-OCT-1999;	99US-0158363P;
13-OCT-1999;	99US-0159293P;
13-OCT-1999;	99US-0159294P;
13-OCT-1999;	99US-0159295P;
14-OCT-1999;	99US-0159329P;
14-OCT-1999;	99US-0159331P;
14-OCT-1999;	99US-0159637P;
21-OCT-1999;	99US-0160771P;
21-OCT-1999;	99US-0160767P;
21-OCT-1999;	99US-0160768P;
21-OCT-1999;	99US-0160770P;
21-OCT-1999;	99US-0160815P;
22-OCT-1999;	99US-0160980P;
22-OCT-1999;	99US-0160981P;
22-OCT-1999;	99US-0160989P;
25-OCT-1999;	99US-0161404P;
25-OCT-1999;	99US-0161405P;
25-OCT-1999;	99US-0161406P;
26-OCT-1999;	99US-0161359P;
26-OCT-1999;	99US-0161360P;
26-OCT-1999;	99US-0161361P;
28-OCT-1999;	99US-0161920P;
28-OCT-1999;	99US-0161922P;
28-OCT-1999;	99US-0161923P;
29-OCT-1999;	99US-0162142P;

Query Match 42.0%: Score 199: DB 3: Length 717:

Query Match  
Post local similarity 77 5%  
pred. No. 5.4e-53:

Best Local Similarity 77.38; Fied: NO: 3.48-35  
Matches 241; Consecutive 0; Mismatches 70

Matches 241; Conservative 0; Mismatches 70

C  
C  
T  
F  
E  
E  
F  
F  
F  
T  
A  
U  
Z  
  
A  
A  
O  
B  
C  
D  
I  
L  
M  
N  
P  
Q  
R  
S  
T  
V

---

1 ATGGCTCGTACGAGCAAACTGCAAGAAAATCACACGG

\_\_\_\_\_

b  
66 ATGGCTCGTACTAAGCAAACAGCTCGTAAGTCTACTGG

[illegible]

61 CTGGCTACCAAGGCGGCGAGGAAATCTGCGCCGACTAC

[illegible]

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

b) 126 CTTGCTACAAAGGCTGCACGTAAGTCTGCACCAACCAAG

121 CGTTACCGTCCCGGAACCGTCGCTCTTCGTGAGATTCCG

\_\_\_\_\_

186 CGTTACCGTCCAGGAAC TGTTCACCTACGTGAAATTCG

**\*\*\*\*\***

101 TTCTTCATCCGTAAACCTTCCTTTCAACGTCTTGTTCC

181 TGGTGAACCGTAAGCTCCIIICAAAGCCTGTTCT

100

246 TTGCTGATCAGGAAGCTCCCCTTCCAGAGGCTAGTTC

Tue Jun 29 08:37:07 2004

QY 241 ACGGATCTGAGATCCAGAGCCATGCGGTGTAGCTCTTCAAGAGCTGCTGAACATAT 300  
Db |||||  
QY 306 ACTGACTTGGCGTTCCAGAGCCATGCTGTGCTTGCACTCCAGGAGGCTGCTGAGCATAC 365  
Db |||||  
QY 301 TTGACGGATCT 311  
Db |||||  
QY 366 CTTGTGGGTCT 376  
Db |||||

Search completed: June 28, 2004, 16:36:37  
Job time : 263.425 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 16:23:14 ; Search time 129 Seconds  
(without alignments)  
357.017 Million cell updates/sec

Title: AAC50492

Perfect score: 852

Sequence: 1 MADEKLPQWGRMSRSG.....GEMSGPFTDGIHILRTE 163

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

- 1: Geneseq1980s.\*
- 2: Geneseq1990s.\*
- 3: Geneseq2000s.\*
- 4: Geneseq2001s.\*
- 5: Geneseq2002s.\*
- 6: Geneseq2003as.\*
- 7: Geneseq2003bs.\*
- 8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	852	100.0	163	2 AAW18312	AAW18312 NIMA-inte
2	852	100.0	163	4 AAG66532	Aag66532 NIMA-inte
3	852	100.0	163	7 ADC72966	Adc72966 Human pet
4	734	86.2	163	2 AAY48377	Aay48377 Human pro
5	705	82.7	131	4 ABG12572	Abg12572 Novel hum
6	642	75.4	259	4 ABG11947	Abg11947 Novel hum
7	580	68.1	195	4 AAU32052	Aau32052 Novel hum
8	456.5	53.6	166	6 ABB66150	Abb66150 Drosophil
9	437.5	51.3	178	6 ABJ25785	Abj25785 Aspergill
10	434.5	51.0	178	6 ABJ26385	Abj26385 Aspergill
11	406	47.7	87	4 AAB87657	Aab87657 Bovine ma
12	373	43.8	70	4 ABB43821	Abb43821 Peptide #
13	373	43.8	70	4 AAM77544	Aam77544 Human bon
14	373	43.8	70	4 AAM77544	Aam77544 Human bon
15	373	43.8	70	4 AAM64786	Aam64786 Human bra
16	373	43.8	70	5 ABG46568	Abg46568 Human pep
17	369.5	43.4	177	3 AAY95876	Aay95876 Candida a
18	369.5	43.4	177	5 AAB73537	Aab73537 Candida a
19	361	42.4	209	7 ADB69989	Adb69989 C. neofo
20	301	35.3	119	3 AAG43433	Aag43433 Arabidops
21	301	35.3	119	3 AAG43433	Aag43433 Arabidops
22	301	35.3	119	3 AAG43433	Aag43433 Arabidops
23	233	27.3	35	5 AAB35511	Aab35511 Human iso
24	215	25.2	39	5 AAB21975	Aab21975 Pinl/huma
25	191	22.4	34	4 AAB74938	Aab74938 Peptidyl

26	186	21.8	155	4	ABG11946	Abg11946 Novel hum
27	179.5	21.1	92	3	AAI90945	Aay90945 Cenarchae
28	179	21.0	114	4	ABG12571	Abg12571 Novel hum
29	172	20.2	31	3	AAB21943	Aab21943 Pinl/huma
30	158	18.5	142	3	AAG11000	Aag11000 Arabidops
31	158	18.5	169	3	AAG10999	Aag10999 Arabidops
32	157.5	18.5	130	4	ABE68428	Abb68428 Drosophil
33	152	17.8	131	2	AAW79203	Aaw79203 Human par
34	152	17.8	131	5	ABO5432	Abb05432 Human par
35	143	16.8	138	3	AAG00864	Aag00864 Human sec
36	138	16.2	92	5	AAO18579	Aao18579 E coli pa
37	137	16.1	333	6	ABP79960	Abp79960 N. gonorr
38	135.5	15.9	330	5	ABP38966	Abp38966 Staphyloc
39	130.5	15.3	1035	4	ABB61120	Abb61120 Drosophil
40	129	15.1	115	3	AAG34786	Aag34786 Arabidops
41	129	15.1	143	3	AAG34784	Aag34784 Arabidops
42	127.5	15.0	320	2	AAW89744	Aaw89744 Staphyloc
43	127.5	15.0	320	6	ABJ18994	Abj18994 Pathogen
44	127.5	15.0	320	6	ABW71095	Abw71095 Staphyloc
45	123	14.4	1082	4	ABB63355	Abb63355 Drosophil

## ALIGNMENTS

RESULT 1  
AAW18312  
ID AAW18312 standard; protein; 163 AA.  
XX  
AC AAW18312;  
XX  
DT 09-DSC-1997 (first entry)  
XX  
DE NIMA-interacting protein Pinl.  
XX  
KW Pinl; protein interacting with NIMA; cell proliferation; mitosis;  
KW peptidyl-propyl cis/trans isomerase; adenocarcinoma; cancer; leukaemia;  
KW psoriasis; pemphigus vulgaris; rheumatoid arthritis;  
KW acute respiratory distress syndrome; septic shock; inflammation; therapy.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Domain 4..43  
FT /label= WW  
FT /note= "Conserved tryptophan domain thought to be mediate  
FT protein-protein interactions"  
FT Peptide 54..69  
FT /note= "putative nuclear localisation signal"  
FT Domain 59..163  
FT /label= PPI  
FT /note= "peptidyl-propyl cis/trans isomerase domain"  
XX  
WO9717986-A1.  
XX  
22-MAY-1997.  
XX  
28-OCT-1996; 96WO-US017334.  
XX  
13-NOV-1995; 95US-00555912.  
XX  
(SALK ) SALK INST BIOLOGICAL STUDIES.  
XX  
PI Hunter T, Lu KP;  
XX  
WPI; 1997-289057/26.  
XX  
N-PSDB; AAT68888.  
XX  
PT Protein, Pinl, interacting with NIMA - used for treating cell  
XX proliferative disorders.  
XX  
Claim 4; Page 49-50; 73pp; English.  
XX

CC This human polypeptide sequence comprises Pin1, an 18 kDa protein that  
 CC has peptidyl-propyl cis/trans isomerase activity, associates with NIMA  
 CC protein kinase, inhibits the mitosis-promoting function of NIMA when  
 CC overexpressed, and induces mitotic arrest and nuclear fragmentation when  
 CC depleted. Its sequence was deduced from a DNA sequence (AA168886)  
 CC identified in an HeLa library using a yeast two-hybrid system. A  
 CC recombinant expression vector comprising the DNA sequence and host cells  
 CC containing the vector are claimed. Methods are also claimed for  
 CC identifying proteins that inhibit the mitosis promoting function of NIMA  
 CC protein kinase and for controlling the growth of a cell by reducing Pin1  
 CC activity or Pin1 expression using an inhibitor, anti-Pin1 antibody,  
 CC antisense nucleotide sequence or ribozyme, or by increasing Pin1 activity  
 CC in the presence of an activator or increasing Pin1 expression using an  
 CC enhancer. This allows treatment of cell proliferation disorders such  
 CC as adenocarcinomas, cancers, psoriasis, pemphigus vulgaris, acute  
 CC respiratory distress syndrome, rheumatoid arthritis, septic shock and  
 CC inflammation

XX Sequence 163 AA;

Query Match 100.0%; Score 852; DB 2; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-83;  
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MADEEKLPGWEKMSRSGRVYFNHITNASQWERPSGSGGKNGGQEPARVCSHL 60  
 DB 1 MADEEKLPGWEKMSRSGRVYFNHITNASQWERPSGSGGKNGGQEPARVCSHL 60  
 QY 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120  
 DB 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120  
 QY 121 DLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGHIILRTE 163  
 DB 121 DLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGHIILRTE 163

RESULT 2

AAAG66532  
 ID AAG66532 standard; protein; 163 AA.

XX AAG66532;

XX 22-OCT-2001 (first entry)

XX Human interferon-alpha induced polypeptide, PIN-1.

XX Human; interferon-alpha induced gene; type I interferon treatment;  
 KW chronic viral hepatitis; relapsing remitting multiple sclerosis;  
 KW neoplastic disease; PIN-1.

XX Homo sapiens.

XX WO200159155-A2.

XX 16-AUG-2001.

XX 09-FEB-2001; 2001WO-GH000578.

XX 11-FEB-2000; 2000GB-00003203.

XX 11-FEB-2000; 2000GB-00003204.

XX 11-FEB-2000; 2000GB-00003205.

XX 11-FEB-2000; 2000GB-00003206.

XX 11-FEB-2000; 2000GB-00003207.

XX 11-FEB-2000; 2000GB-00003208.

XX 11-FEB-2000; 2000GB-00003210.

XX 11-FEB-2000; 2000GB-00003212.

XX 11-FEB-2000; 2000GB-00003213.

XX 11-FEB-2000; 2000GB-00003215.

XX 11-FEB-2000; 2000GB-00003216.

PR 11-FEB-2000; 2000GB-00003222.  
 PR 17-FEB-2000; 2000GB-00003768.

XX (PHAR-) PHARMA PACIFIC PTY LTD.

XX Meritet J, Dron M, Tovey MG;

XX WPI: 2001-483570/52.

DR N-PSDB; AAH76459.

XX Predicting responsiveness of a patient to treatment with a type I

PT interferon comprising determining the level of induced proteins after

PT treatment with a type I interferon.

XX Claim 1; Page 61-62; 133pp; English.

CC The invention relates to a method for predicting responsiveness of a

CC patient to treatment with a type I interferon. The method comprises

CC determining the level of one or more proteins with a 646, 164, 126, 598,

CC 98, 177, 761, 361, 941, 657, 817, 429, 473, 399, 285 or 303 amino acid

CC sequence fully defined in the specification after treatment with a type I

CC interferon. The method allows a physician to determine whether a patient

CC suffering from chronic viral hepatitis, neoplastic disease or relapsing

CC remitting multiple sclerosis will respond favourably to type I interferon

CC treatment via oromucosal administration. The present sequence is one of

CC the polypeptides listed above that may be used in the method

XX Sequence 163 AA;

Query Match 100.0%; Score 852; DB 4; Length 163;

Best Local Similarity 100.0%; Pred. No. 3.5e-83;

Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MADEEKLPGWEKMSRSGRVYFNHITNASQWERPSGSGGKNGGQEPARVCSHL 60

DB 1 MADEEKLPGWEKMSRSGRVYFNHITNASQWERPSGSGGKNGGQEPARVCSHL 60

QY 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120

DB 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120

QY 121 DLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGHIILRTE 163

DB 121 DLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGHIILRTE 163

RESULT 3

ADCT72966

ID ADCT72966 standard; protein; 163 AA.

XX ADCT72966;

XX 01-JAN-2004 (first entry)

XX Human petidyl polyl isomerase protein.

XX cytostatic; gynaecological; antipsoriatic; antimicrobial; antibacterial;

XX antiinflammatory; anti diarrhoeic; antipyretic; antirheumatic;

XX immunosuppressive; gastrointestinal; antitubercular; tuberculostatic;

XX fungicide; Pini; petidyl polyl isomerase; PPIase; cancer;

XX fibrotic disorder; non-neoplastic growth; benign hypertrophy;

XX endometriosis; psoriasis; infectious disease; fungal;

XX bacterial infection; pneumonia; diarrhoea; dysentery; anthrax;

XX rheumatic fever; toxic shock syndrome; mastoiditis; meningitis;

XX gonorrhoea; typhoid fever; gastroenteritis; brucellosis; cholera;

XX bubonic plague; tetanus; tuberculosis; Lyme disease; human; enzyme.

XX Homo sapiens.

XX WO2003074002-A2.

XX 12-SEP-2003.

PF 03-MAR-2003; 2003WO-US006518.  
 XX  
 PR 01-MAR-2002; 2002US-0360799P.  
 XX  
 PA (PINT-) PINTEX PHARM INC.  
 XX  
 PI Suto RK, Sowadski J;  
 XX  
 XX WPI; 2003-779008/73.  
 XX  
 XX Treating peptidyl prolyl isomerase Pin1 associated disorder e.g. cancer,  
 PT infectious diseases and psoriasis comprises administering Pin1 activity  
 PT inhibitors.  
 PT  
 XX Disclosure; SEQ ID NO 1; 4lpp; English.  
 XX  
 XX The invention relates to a novel method of treating a Pin1 (petidyl polyI  
 CC isomerase; ppiase)-associated disorder comprising administration of Pin1  
 CC activity inhibitors. The method of the invention may be useful for  
 CC treating disorders involving mitosis and cell proliferation, particularly  
 CC cancers, fibrotic disorders and non-neoplastic growths including benign  
 CC hyper trophy, endometriosis and psoriasis, as well as infectious diseases  
 CC including fungal and bacterial infections such as pneumonia, diarrhoea,  
 CC dysentery, anthrax, rheumatic fever, toxic shock syndrome, mastoiditis,  
 CC meningitis, gonorrhoea, typhoid fever, gastroenteritis, brucellosis,  
 CC cholera, bubonic plague, tetanus, tuberculosis and Lyme disease. The  
 CC current sequence is that of the human Pin1 protein of the invention.  
 XX  
 XX SQ Sequence 163 AA;  
 Query Match 100.0%; Score 852; DB 7; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-83;  
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MADEKLPQGWKEMSRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60  
 Db 1 MADEKLPQGWKEMSRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60  
 Qy 61 LVKHSQRRPSSWRQEKITRKEALELINGYIOKIKSGEDEFSLASQFSDCSAKARG 120  
 Db 61 LVKHSQRRPSSWRQEKITRKEALELINGYIOKIKSGEDEFSLASQFSDCSAKARG 120  
 Qy 121 DLGAFSRGQMKPPEDASFALRTGEMSGPVFTDSGIHILRTE 163  
 Db 121 DLGAFSRGQMKPPEDASFALRTGEMSGPVFTDSGIHILRTE 163  
 RESULT 4  
 AAY48377  
 XX AAY48377 standard; protein; 163 AA.  
 XX  
 AC AAY48377;  
 XX  
 XX 08-DEC-1999 (first entry)  
 DT  
 XX Human prostate cancer-associated protein 74.  
 DE  
 XX Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;  
 KW cancer; tissue specificity; human.  
 XX  
 XX Homo sapiens.  
 OS  
 XX DE19811194-A1.  
 PN  
 XX 16-SEP-1999.  
 PD  
 XX 10-MAR-1998; 98DE-01011194.  
 PF  
 XX 10-MAR-1998; 98DE-01011194.  
 PR  
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 PA  
 XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;

XX WPI; 1998-519629/44.  
 DR N-PSDB; AAZ33510.  
 XX  
 PT New nucleic acid expressed at high level in normal prostatic tissue and  
 PT encoded polypeptides, used to treat cancer and screen for therapeutic  
 PT agents.  
 XX  
 PS Claim 25; 152; 194pp; German.  
 XX  
 XX This invention describes novel nucleic acid sequences (A) that are  
 CC expressed at high level in normal prostatic tissue. Polypeptides (I)  
 CC encoded by (A) are used: (a) for identifying agents for treatment of  
 CC prostatic cancer and (b) for therapy of prostate cancer, optionally where  
 CC expressed by gene therapy methods. (A) is also used to isolate full-  
 CC length genes (for gene therapy) and for recombinant production of (I),  
 CC which can be used to raise specific antibodies. (A) are identified by  
 CC assembly of ESTs (expressed sequence tags) before these are analyzed for  
 CC expression pattern (tissue specificity). This approach eliminates many of  
 CC the false results, as regards tissue specificity, associated with known  
 CC methods that use single (usually short) ESTs. AAY48304-Y48456 represent  
 CC peptides encoded by the expressed sequence tags described in the method  
 CC of the invention  
 XX  
 XX SQ Sequence 163 AA;  
 Query Match 86.2%; Score 734; DB 2; Length 163;  
 Best Local Similarity 97.2%; Pred. No. 1.7e-70;  
 Matches 140; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 MADEKLPQGWKEMSRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60  
 Db 1 MADEKLPQGWKEMSRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60  
 Qy 61 LVKHSQRRPSSWRQEKITRKEALELINGYIOKIKSGEDEFSLASQFSDCSAKARG 120  
 Db 61 LVKHSQRRPSSWRQEKITRKEALELINGYIOKIKSGEDEFSLASQFSDCSAKARG 120  
 Qy 121 DLGAFSRGQMKPPEDASFALRTG 144  
 Db 121 DLGAFSRGQMKPPEDASFALRTG 144  
 RESULT 5  
 ABG12572  
 ID ABG12572 standard; protein; 191 AA.  
 XX  
 AC ABG12572;  
 XX  
 XX 18-FEB-2002 (first entry)  
 DT  
 XX Novel human diagnostic protein #12563.  
 DE  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200175067-A2.  
 PN  
 XX 11-OCT-2001.  
 PD  
 XX 30-MAR-2001; 2001WO-US008631.  
 PF  
 XX 31-MAR-2000; 2000US-00540217.  
 PR  
 XX 23-AUG-2000; 2000US-00649167.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 XX WPI; 2001-639362/73.  
 DR  
 DR N-PSDB; AAS76759.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX Claim 20; SEQ ID NO 42931; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 191 AA;

Query Match 82.7%; Score 705; DB 4; Length 191;  
Best Local Similarity 82.6%; Pred. No. 2.8e-67;  
Matches 142; Conservative 2; Mismatches 8; Indels 20; Gaps 3;

QY 9 PGWEXMSRSSG-----RVYFNFHITNASQWRPSSGSGKNGOG 50  
DB 1 PGWEXMSRSSGVTQEAIPDAAIPDAAKRVYFNFHITNASQWRPSSGSGKNGOG 60

QY 51 EPARVRCSHLLVKHSQSRPSSWRQEKITRTKEALELINGYIQTKSGEEDFESLASOF 110  
DB 61 EPARVRCSHLLVKHSQSRPSSWRQEKITRTKEALELINGYIQTKSGEEDFESLASOF 120

QY 111 SDCCSAKARGDLGAFSGQWQKPF-EDASPALRT-GEMSGPVFTDSGIHIL 160  
DB 121 SDCCSAKARGDLGFSRGQWQKPFRTPRFALRTGGDERGFCFTDSGIHIL 172

RESULT 6  
ABG11947  
ID ABG11947 standard; protein; 259 AA.  
XX AC ABG11947;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #11938.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN W0200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX PR 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.  
PI Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS76134.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX Claim 20; SEQ ID NO 42306; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 259 AA;

Query Match 75.4%; Score 642; DB 4; Length 259;  
Best Local Similarity 76.4%; Pred. No. 2.5e-60;  
Matches 133; Conservative 5; Mismatches 14; Indels 22; Gaps 5;

QY 9 PGWEXMSRSS-----GRVYFNFHITNASQWRPSSGSGKNGOG 50  
DB 1 PGWEXMSRSSVVTQEAIPDAAIPDAAKRVYFNFHITNASQWRPSSGSGKNGOG 60

QY 51 EPARVRCSHLLVKHSQSRPSSWRQEKITRTKEALELINGYIQTKSGEEDFESLASOF 110  
DB 61 EPARVRCSHLLVKHSQSRPSSWRQEKITRTKEALELINGYIQTKSGEEDFESLASOF 120

QY 111 SDCCSAKARGDLGAF-SRGQWQKPF-EDASPALRT-GEMSGPVFTDSGIHIL 160  
DB 121 SDCCSAKARGDLGAFQKVRQKPFRTPRFALDGGKXAGCFTDSGIHIL 174

RESULT 7  
AAU32052  
ID AAU32052 standard; protein; 195 AA.  
XX AC AAU32052;  
XX DT 18-DEC-2001 (first entry)  
XX DE Novel human secreted protein #2543.  
XX KW Human; vaccination; gene therapy; nutritional supplement;  
XX stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
XX immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
XX OS Homo sapiens.  
XX PN W0200179449-A2.

```

PD XX 25-OCT-2001.
PF XX
XX 16-APR-2001; 2001WO-US008656.
XX PR
XX 18-APR-2000; 2000US-00552929.
XX PR
XX 26-JAN-2001; 2001US-00770160.
XX PR
XX (HYSE-) HYSEQ INC.
XX PA
XX Tang YT, Liu C, Drmanac RT;
XX PI
XX WPI; 2001-611725/70.
XX DR
XX Nucleic acids encoding a range of human polypeptides, useful in genetic
XX PT vaccination, testing and therapy.
XX PT
XX Claim 20; Page 548-549; 765pp; English.
XX PS
XX The invention relates to novel human secreted polypeptides. The
XX CC polypeptides and antibodies to the polypeptides are useful for
XX CC determining the presence of or predisposition to a disease associated
XX CC with altered levels of polypeptide. The polypeptides are also useful for
XX CC identifying agents (agonists and antagonists) that bind to them. Cells
XX CC expressing the proteins are useful for identifying a therapeutic agent
XX CC for use in treatment of a pathology related to aberrant expression or
XX CC physiological interactions of the polypeptide. Vectors comprising the
XX CC nucleic acids encoding the polypeptides and cells genetically engineered
XX CC to express them are also useful for producing the proteins. The proteins
XX CC are useful in genetic vaccination, testing and therapy, and can be used
XX CC as nutritional supplements. They may be used to increase stem cell
XX CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
XX CC and/or nerve tissue growth or regeneration; immune suppression and/or
XX CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
XX CC AAU29510-AAU33304 represent the amino acid sequences of novel human
XX CC secreted proteins of the invention
XX CC
XX SQ Sequence 195 AA;
XX
Query Match 68.1%; Score 580; DB 4; Length 195;
Best Local Similarity 72.2%; Pred. No. 7.9e-54;
Matches 127; Conservative 7; Mismatches 18; Indels 24; Gaps 7;
QY 9 PGWKKRMRSS-----GRVYFNHITNASQWRPSSGSGGKNGQG 50
Db 1 PGWKKRMRSSVVNTQEQALPTAIPDAKGRVYFNHITNASQWRPSSGSGGKNGQG 60
QY 51 EPARVRCSHLVKHSQSRPPSSWRQEKITRTKKEAL-ELING-YTQTKSGEEDFESLAS 108
Db 61 EPVAVRCSHLLVKHSQSRPPSSWRQEKITRTKGGGPGSLINGLHPEDSSGGEEDFESLAS 120
QY 109 QPSDC-SSAKARGDYGAF-SRQOMQKPF-EDASFALRT-GENMSGPVFTDSDGIHILL 160
Db 121 QPSDCSSAKARGDLGAFQQKVRLOKPFILKTRFALRTGGDRGRCFTDSDGIHILL 176
RESULT 8
ABE66150
ID ABB66150 standard; protein; 166 AA.
XX AC
XX ABB66150;
XX DT
XX 26-MAR-2002 (first entry)
XX DE
XX Drosophila melanogaster polypeptide SEQ ID NO 25242.
XX KW
XX Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS
XX Drosophila melanogaster.
XX XX
XX WO200171042-A2.
XX PN
XX 27-SEP-2001.
XX PD

```

```

XX 23-MAR-2001; 2001WO-US009231.
XX PF
XX 23-MAR-2000; 2000US-0191637P.
XX PR
XX 11-JUL-2000; 2000US-00614150.
XX PR
XX (PEKE ) PE CORP NY.
XX PA
XX Venter JC, Adams M, Li PWD, Myers EW;
XX PI
XX WPI; 2001-656860/75.
XX DR
XX N-PSDB; ABL10253.
XX XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX PT
XX Disclosure; SEQ ID NO 25242; 21pp + Sequence Listing; English.
XX PS
XX The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
XX CC AB872072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX CC
XX SQ Sequence 166 AA;
XX
Query Match 53.6%; Score 456.5; DB 4; Length 166;
Best Local Similarity 56.7%; Pred. No. 1.2e-40;
Matches 93; Conservative 16; Mismatches 52; Indels 3; Gaps 2;
QY 1 MADEKLPDGGWEKMRSSGRVYFNHITNASQWRPSS--GNSSSGGKNGQGE-PARVRC 57
Db 1 MPDAEQLPDGGWEKTSRSTGMSYVLYNMYTKESQWDQTEPAKAGGSAGGDAPDEVHC 60
QY 58 SHLLVKSQSRPPSSWRQEKITRTKKEALELINGYIOKIKSGEEDFESLASQFSDCSSAK 117
Db 61 LHLLVKKHSGSRPPSSWRREANITRTKKEAQLLLLVYRNKIVQOEATFDELARSYSDCSSAK 120
QY 118 ARGDLGAFSRQOMQKPFEDASFALRTGEMSGPVFTDSDGIHILLR 161
Db 121 RGGDLGKFGRCQMQAPEADAAFKUNVQLSGLVDSGSLHILLR 164
RESULT 9
ABJ25785
ID ABJ25785 standard; protein; 178 AA.
XX AC
XX ABJ25785;
XX DT
XX 16-APR-2003 (first entry)
XX DE
XX Aspergillus fumigatus essential gene protein #443.
XX KW
XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
XX KW cancer; contamination; biofilm; antibody; immune response.
XX OS
XX Aspergillus fumigatus.
XX XX
XX WO200286090-A2.
XX PN
XX 31-OCT-2002.
XX PD
XX 23-APR-2002; 2002WO-US013142.
XX PF
XX 23-APR-2001; 2001US-0285697P.
XX PR
XX 27-APR-2001; 2001US-0287066P.
XX PR
XX 05-JUN-2001; 2001US-0295890P.
XX PD

```

```

PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX
XX New purified or isolated nucleic acids of essential genes of Aspergillus
XX fumigatus, useful for treating or preventing infections by A. fumigatus,
XX or for treating a non-infectious disease in a subject e.g. cancer.
XX
XX Disclosure; Page; 175pp; English.
XX
XX The invention relates to novel purified or isolated nucleic acids of
XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of
XX the invention are used to treat or prevent infections by a pathogenic
XX organism such as A. fumigatus, to treat a non-infectious disease in a
XX subject (e.g. cancer), to prevent or inhibit formation of an object
XX by A. fumigatus, or to prevent or inhibit formation on a surface of a
XX biofilm comprising A. fumigatus. The polynucleotides are useful for
XX expressing recombinant protein for characterisation, screening or
XX potential orthologous essential or virulence genes, for selecting and
XX making oligomers for attachment to a nucleic acid array for examination
XX of expression patterns, for raising anti-protein antibodies, as an
XX antigen to raise anti-DNA antibodies or to elicit another immune
XX response, and for identifying polynucleotides encoding the other protein
XX with which binding occurs or to identify inhibitors of the binding
XX determine levels of the protein in biological fluids, as a marker for
XX host tissues in which pathogenic organism invade or reside, and to
XX isolate correlative receptors or ligands in the case of virulence
XX factors. This sequence represents a protein of one of the essential genes
XX of Aspergillus fumigatus of the invention
XX
XX Sequence 178 AA;
XX
Query Match 51.3%; Score 437.5; DB 6; Length 178;
Best Local Similarity 51.5%; Pred. No. 1.4e-38;
Matches 88; Conservative 26; Mismatches 40; Indels 17; Gaps 3;

Qy 7 LPFGWEKRMRSRSGRVVYFNHITNASQWERPSG-----NSSGGK---NQQG 50
Db 9 LPAGWEVHRHNSKNLPYYFNPSTKESRWEPPSGTDTLLKVMYANYHSAPAGRDPGTAQG 68

Qy 51 EPARVRCSHLLVKSQSRPSSWROEKITRTKEAEALINGYIQIKSGEDFESLASQF 110
Db 69 E-GKIRCSHLLIKHRDSRPPSSWREAEITRSKEAEIILRGHEQIRSGEVSGLGDIAYSE 127

Qy 111 SDCCSSAKARGDLGAFSGQMKPFEDASFALRTGEMSGPVFTDSGIHILR 161
Db 128 SDCCSARKKGDGFGRGEMQKEFEDAAAFALQPGQVSGIVETASGVHLIER 178

RESULT 10
ABJ26385
ID ABJ26385 standard; protein; 178 AA.
XX
XX AC ABJ26385;
XX
XX 16-APR-2003 (first entry)
XX
XX Aspergillus fumigatus essential gene protein #1043.
XX
XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
XX cancer; contamination; biofilm; antibody; immune response.

```

```

XX OS Aspergillus fumigatus.
XX
XX PN WO200286090-A2.
XX
XX PD 31-OCT-2002.
XX
XX PF 23-APR-2002; 2002WO-US013142.
XX
XX PR 23-APR-2001; 2001US-0285697P.
XX 27-APR-2001; 2001US-0287066P.
XX 05-JUN-2001; 2001US-0295890P.
XX 09-JUL-2001; 2001US-0303899P.
XX 31-AUG-2001; 2001US-0316362P.
XX
XX PA (ELIT-) ELITRA PHARM INC.
XX
XX PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX
XX DR New purified or isolated nucleic acids of essential genes of Aspergillus
XX fumigatus, useful for treating or preventing infections by A. fumigatus,
XX or for treating a non-infectious disease in a subject e.g. cancer.
XX
XX PS Disclosure; Page; 175pp; English.
XX
XX CC The invention relates to novel purified or isolated nucleic acids of
XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of
XX the invention are used to treat or prevent infections by a pathogenic
XX organism such as A. fumigatus, to treat a non-infectious disease in a
XX subject (e.g. cancer), to prevent or inhibit formation of an object
XX by A. fumigatus, or to prevent or inhibit formation on a surface of a
XX biofilm comprising A. fumigatus. The polynucleotides are useful for
XX expressing recombinant protein for characterisation, screening or
XX therapeutic use, as markers for host tissues in which the pathogenic
XX organism invade or reside, for comparing with the DNA sequence of A.
XX fumigatus to identify duplicated genes or paralogues having the same or
XX similar biochemical activity and/or function, for comparing with DNA
XX sequences of other related or distant pathogenic organisms to identify
XX potential orthologous essential or virulence genes, for selecting and
XX making oligomers for attachment to a nucleic acid array for examination
XX of expression patterns, for raising anti-protein antibodies, as an
XX antigen to raise anti-DNA antibodies or to elicit another immune
XX response, and for identifying polynucleotides encoding the other protein
XX with which binding occurs or to identify inhibitors of the binding
XX interaction. The polypeptides may be used to raise antibodies or to
XX elicit immune response, as a reagent in assays designed to quantitatively
XX determine levels of the protein in biological fluids, as a marker for
XX host tissues in which pathogenic organism invade or reside, and to
XX isolate correlative receptors or ligands in the case of virulence
XX factors. This sequence represents a protein of one of the essential genes
XX of Aspergillus fumigatus of the invention
XX
XX Sequence 178 AA;
XX
Query Match 51.0%; Score 434.5; DB 6; Length 178;
Best Local Similarity 50.6%; Pred. No. 3e-38;
Matches 88; Conservative 26; Mismatches 43; Indels 17; Gaps 3;

Qy 4 EEKLPPGWEKRMRSRSGRVVYFNHITNASQWERPSG-----NSSGGK---N 47
Db 6 ETGLPAGWEVHRHNSKNLPYYFNPSTKESRWEPPSGTDTLLKVMYANYHSAPAGRDPGT 65

Qy 48 GQGEPAVRCSHLLVKSQSRPSSWROEKITRTKEAEALINGYIQIKSGEDFESLA 107
Db 66 AQGE-GKIRCSHLLIKHRDSRPPSSWREAEITRSKEAEIILRGHEQIRSGEVSGLGDI 124

Qy 108 SQFSDCCSARKARGDLGAFSGQMKPFEDASFALRTGEMSGPVFTDSGIHILR 161
Db 125 VSESDCCSARKKGDGFGRGEMQKEFEDAAAFALQPGQVSGIVETASGVHLIER 178

```



```
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48897/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX Claim 27; SEQ ID NO 37998; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP:
XX see AAI31315-AA157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX
XX Sequence 70 AA;
XX
XX Query Match 43.8%; Score 373; DB 4; Length 70;
XX Best Local Similarity 100.0%; Pred. No. 3.4e-32;
XX Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 21 RVYFNHITNASQWERPSGNSGGKNGGQGEPAVRCSHLLVYKHQSRRPSSWRQEKITR 80
DB 1 RVYFNHITNASQWERPSGNSGGKNGGQGEPAVRCSHLLVYKHQSRRPSSWRQEKITR 60
XX
QY 81 TKEEALIELIN 90
DB 61 TKEEALIELIN 70
XX
XX RESULT 14
XX AAM77544
XX ID AAM77544 standard; protein; 70 AA.
XX AC AAM77544;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 37850.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000668.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX
```

```
XX Example 4; SEQ ID NO 37850; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX
XX Sequence 70 AA;
XX
XX Query Match 43.8%; Score 373; DB 4; Length 70;
XX Best Local Similarity 100.0%; Pred. No. 3.4e-32;
XX Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 21 RVYFNHITNASQWERPSGNSGGKNGGQGEPAVRCSHLLVYKHQSRRPSSWRQEKITR 80
DB 1 RVYFNHITNASQWERPSGNSGGKNGGQGEPAVRCSHLLVYKHQSRRPSSWRQEKITR 60
XX
QY 81 TKEEALIELIN 90
DB 61 TKEEALIELIN 70
XX
XX RESULT 15
XX AAM64786
XX ID AAM64786 standard; protein; 70 AA.
XX AC AAM64786;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 36891.
XX
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000667.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
XX Example 4; SEQ ID NO 36891; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention
XX
```



Tue Jun 29 08:37:06 2004

SQ Sequence 70 AA;  
Query Match 43.8%; Score 373; DB 4; Length 70;  
Best Local Similarity 100.0%; Pred.No. 3.4e-32;  
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 21 RYVYFNHITNASQWERPSGNSGGKNGQGEPAFVRCSHLLVKHQSRRPSSWQEKITR 80  
Db 1 RYVYFNHITNASQWERPSGNSGGKNGQGEPAFVRCSHLLVKHQSRRPSSWQEKITR 60  
Qy 81 TKEALELIN 90  
Db 61 TKEALELIN 70

Search completed: June 28, 2004, 18:36:17  
Job time : 132 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 18:34:06 ; Search time 30 Seconds  
(without alignments)  
280.501 Million cell updates/sec

Title: AAC50492

Perfect score: 852

Sequence: 1 MADEEKLPPGWEKMRSSG.....GEMSGPVFTDSGHIILRTS 163

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pap.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pap.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pap.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pap.\*  
5: /cgn2\_6/ptodata/2/iaa/6CTUS COMB.pap.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	852	100.0	163	2	US-09-066-074-2
2	852	100.0	163	2	US-08-555-912A-2
3	852	100.0	163	3	US-09-208-804-4
4	852	100.0	163	3	US-08-801-743-4
5	852	100.0	163	4	US-09-275-900-2
6	369.5	43.4	177	4	US-09-507-242-2
7	215	25.2	39	2	US-09-066-074-8
8	215	25.2	39	2	US-08-555-912A-8
9	215	25.2	39	4	US-09-252-404A-33
10	215	25.2	39	4	US-09-275-900-8
11	179.5	21.1	92	4	US-09-408-020-68
12	172	20.2	31	4	US-09-252-404A-1
13	152	17.8	131	3	US-09-208-804-1
14	152	17.8	131	3	US-08-801-743-1
15	152	17.8	131	4	US-09-621-976-7038
16	152	17.8	131	4	US-09-976-594-617
17	138.5	16.3	698	4	US-09-252-991A-30206
18	138	16.2	93	3	US-09-208-804-3
19	138	16.2	93	3	US-08-801-743-3
20	135.5	15.9	330	4	US-09-134-001C-3811
21	135	15.8	110	4	US-09-489-039A-8030
22	130.5	15.3	111	4	US-09-252-991A-20022
23	129	15.1	105	4	US-09-489-039A-8015
24	127.5	15.0	320	4	US-08-956-171E-5192
25	118	13.8	443	4	US-09-328-352-7725
26	117	13.7	369	4	US-09-540-236-3738
27	116	13.6	735	3	US-08-539-205A-2

28	116	13.6	735	4	US-09-392-163A-2	Sequence 2, Appli
29	115.5	13.6	852	2	US-09-070-060-3	Sequence 3, Appli
30	115.5	13.6	852	3	US-09-357-746-3	Sequence 3, Appli
31	115.5	13.6	854	2	US-09-070-060-4	Sequence 4, Appli
32	115.5	13.6	854	3	US-09-357-746-4	Sequence 4, Appli
33	114	13.4	97	4	US-09-252-991A-20492	Sequence 20492, A
34	114	13.4	906	3	US-08-630-916A-48	Sequence 48, Appl
35	113	13.3	106	4	US-09-328-352-5211	Sequence 5211, Ap
36	109.5	12.9	38	3	US-08-630-916A-36	Sequence 36, Appl
37	108.5	12.7	683	3	US-08-630-916A-46	Sequence 46, Appl
38	107	12.6	299	4	US-09-336-115C-24	Sequence 24, Appl
39	105.5	12.4	38	3	US-08-630-916A-32	Sequence 32, Appl
40	105.5	12.4	446	4	US-09-489-039A-12283	Sequence 12283, A
41	105	12.3	834	3	US-08-539-205A-6	Sequence 6, Appli
42	105	12.3	834	4	US-09-392-163A-6	Sequence 6, Appli
43	103	12.1	927	3	US-08-895-601-6	Sequence 6, Appli
44	102	12.0	274	4	US-09-252-991A-20389	Sequence 20389, A
45	101	11.9	766	3	US-08-539-205A-4	Sequence 4, Appli

#### ALIGNMENTS

#### RESULT 1

US-09-066-074-2  
; Sequence 2, Application US/09066074  
; Patent No. 5952467  
; GENERAL INFORMATION:  
; APPLICANT: Hunter et al., Tony  
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/066,074  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/555,912  
; FILING DATE: 13-NOV-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07251/011001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 163 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-066-074-2

Query Match 100.0%; Score 852; DB 2; Length 163;

Best Local Similarity 100.0%; Pred. No. 8e-85;

Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MADEEKLPPGWEKMRSSGGRVYVFNHTNASQWERPSGNSGKNGQGEFARVRCSHL 60

Db 1 MADEEKLPPGWEKMRSSGGRVYVFNHTNASQWERPSGNSGKNGQGEFARVRCSHL 60

QY 61 LVKHSQSRPPSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120  
 Db 61 LVKHSQSRPPSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120  
 QY 121 DLGAFSRGQMQKPFEDASFALRTGEMSGPVFTDSDGIHILRTE 163  
 Db 121 DLGAFSRGQMQKPFEDASFALRTGEMSGPVFTDSDGIHILRTE 163

## RESULT 2

US-08-555-912A-2  
 ; Sequence 2, Application US/08555912A  
 ; Patent No. 5972697  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hunter et al., Tony  
 ; TITLE OF INVENTION: NIMA INTERACTING PROTEINS  
 ; NUMBER OF SEQUENCES: 14  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 4225 Executive Square, Suite 1400  
 ; CITY: La Jolla  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/555,912A  
 ; FILING DATE: 13-NOV-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Haile, Lisa A.  
 ; REGISTRATION NUMBER: 38,347  
 ; REFERENCE/DOCKET NUMBER: 07251/011001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619/678-5070  
 ; TELEFAX: 619/678-5099  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 163 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-555-912A-2

Query Match 100.0%; Score 852; DB 2; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 8e-85;  
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGGGEPAVRVCSHL 60  
 Db 1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGGGEPAVRVCSHL 60  
 QY 61 LVKHSQSRPPSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120  
 Db 61 LVKHSQSRPPSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120  
 QY 121 DLGAFSRGQMQKPFEDASFALRTGEMSGPVFTDSDGIHILRTE 163  
 Db 121 DLGAFSRGQMQKPFEDASFALRTGEMSGPVFTDSDGIHILRTE 163

## RESULT 3

US-09-208-804-4  
 ; Sequence 4, Application US/09208804  
 ; Patent No. 6030826  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Au-Young, Janice  
 ; TITLE OF INVENTION: NOVEL HUMAN PARVULIN-LIKE PROTEIN  
 ; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: Fast-SEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/208,804  
 ; FILING DATE: Herewith  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION NUMBER: 08/801,743  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J.  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0217 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-855-0555  
 ; TELEFAX: 415-845-4166  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 163 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: GenBank  
 ; CLONE: 1332710  
 ; US-09-208-804-4

Query Match 100.0%; Score 852; DB 3; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 8e-85;  
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGGGEPAVRVCSHL 60  
 Db 1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGGGEPAVRVCSHL 60  
 QY 61 LVKHSQSRPPSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120  
 Db 61 LVKHSQSRPPSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120  
 QY 121 DLGAFSRGQMQKPFEDASFALRTGEMSGPVFTDSDGIHILRTE 163  
 Db 121 DLGAFSRGQMQKPFEDASFALRTGEMSGPVFTDSDGIHILRTE 163

## RESULT 4

US-08-801-743-4  
 ; Sequence 4, Application US/08801743  
 ; Patent No. 6037164  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Au-Young, Janice  
 ; TITLE OF INVENTION: NOVEL HUMAN PARVULIN-LIKE PROTEIN  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible

```

; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,743
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0217 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 163 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1332710
; US-08-801-743-4

Query Match      100.0%; Score 852; DB 3; Length 163;
Best Local Similarity 100.0%; Pred. No. 8e-85;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPARVRCSHL 60
Db      1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPARVRCSHL 60
QY      61 LVKHSQRRPSSNRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120
Db      61 LVKHSQRRPSSNRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120
QY      121 DLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGHIILRTE 163
Db      121 DLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGHIILRTE 163

RESULT 5
US-09-275-900-2
; Sequence 2, Application US/09275900
; Patent No. 6596848
; GENERAL INFORMATION:
; APPLICANT: Hunter et al., Tony
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/275,900
; FILING DATE: 24-Mar-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/555,912
; FILING DATE: 13-NOV-1995
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07251/011001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 163 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-275-900-2

Query Match      100.0%; Score 852; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 8e-85;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPARVRCSHL 60
Db      1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPARVRCSHL 60
QY      61 LVKHSQRRPSSNRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120
Db      61 LVKHSQRRPSSNRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120
QY      121 DLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGHIILRTE 163
Db      121 DLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGHIILRTE 163

RESULT 6
US-09-507-242-2
; Sequence 2, Application US/09507242
; Patent No. 6537753
; GENERAL INFORMATION:
; APPLICANT: HANES, Steven D.
; APPLICANT: DEVASAHAYAM, Gina
; APPLICANT: CHATURVEDI, Vishnu
; TITLE OF INVENTION: CAESSE: A CANDIDA ALBICANS GENE, METHODS FOR MAKING AND USING, AND TARGETING IT OR ITS EXPRESSION PRODUCTS FOR ANTIFUNGAL APPLICATIONS
; FILE REFERENCE: 454311-2200.1
; CURRENT APPLICATION NUMBER: US/09/507,242
; CURRENT FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 60/121,246
; PRIOR FILING DATE: 1999-02-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-507-242-2

Query Match      43.4%; Score 369.5; DB 4; Length 177;
Best Local Similarity 43.2%; Pred. No. 2.4e-32;
Matches 73; Conservative 34; Mismatches 49; Indels 13; Gaps 2;

QY      7 LPFGWEKMRSSGRVYFNHITNASQWERPSGN-----SSSGKNGQGEPAR 54
Db      8 LPNWTIRVSRSHNKVEYFLNQSTNESSWDPPTYGTDKEVLNAYIAKFKNGYKPLVNEGQ 67
QY      55 VRCSHLIVKHSQRRPSSNRQ-EKIRTKEEALELINGYIQIKSGEEDFESLASQFSDC 113
Db      68 VRVSHLLIKNNQSRKPKSKWSDGIGIRTRDESQILKKGHLRILSGEVLKSELANTESDC 127
QY      114 SSAKARGDLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGHIILRT 162
Db      128 SSHDRGDLGFFSKGQMQPPFEAAFNHVGVSNIETNSGVHILQRT 176

```

```
RESULT 7
US-09-066-074-8
; Sequence 8, Application US/09066074
; Patent No. 5952467
; GENERAL INFORMATION:
; APPLICANT: Hunter et al., Tony
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,074
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/555,912
; FILING DATE: 13-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07251/011001
; TELEPHONE: 619/678-5099
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: PIN1/HUMAN
; US-09-066-074-8

Query Match 25.2%; Score 215; DB 2; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.9e-16;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EKLPGWEKMRSSGRVYFNHITNASQWERPSGNSS 43
DB 1 EKLPGWEKMRSSGRVYFNHITNASQWERPSGNSS 39

RESULT 9
US-09-252-404A-33
; Sequence 33, Application US/09252404A
; Patent No. 6495376
; GENERAL INFORMATION:
; APPLICANT: Kun Ping Lu
; APPLICANT: Xiao Zhen Zhou
; TITLE OF INVENTION: Methods and Compositions for Regulating
; TITLE OF INVENTION: Protein-Protein Interactions
; FILE REFERENCE: 1440.1034-000
; CURRENT APPLICATION NUMBER: US/09/252,404A
; CURRENT FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; TYPE: PRT
; ORGANISM: Homo sapien
; US-09-252-404A-33

Query Match 25.2%; Score 215; DB 4; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.9e-16;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EKLPGWEKMRSSGRVYFNHITNASQWERPSGNSS 43
DB 1 EKLPGWEKMRSSGRVYFNHITNASQWERPSGNSS 39

RESULT 10
US-09-275-900-8
; Sequence 8, Application US/09275900
; Patent No. 6596848
; GENERAL INFORMATION:
; APPLICANT: Hunter et al., Tony
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; US-09-275-900-8
```

STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/275,900  
FILING DATE: 24-Mar-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/555,912  
FILING DATE: 13-NOV-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07251/011001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
IMMEDIATE SOURCE:  
CLONE: PIN1/HUMAN  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-09-275-900-8

Query Match 25.2%; Score 215; DB 4; Length 39;  
Best Local Similarity 100.0%; Pred. No. 1.9e-16;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EKLPPEKMSRSGRVVYFNHITNASQWERPSGSS 43  
DB 1 EKLPPEKMSRSGRVVYFNHITNASQWERPSGSS 39

RESULT 11  
US-09-408-020-68  
Sequence 68, Application US/09408020  
Patent No. 663237  
GENERAL INFORMATION:  
APPLICANT: Swanson, Ronald V.  
APPLICANT: Feldman, Robert A.  
APPLICANT: Schleper, Christa  
TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
FILE REFERENCE: DORP.002A  
CURRENT APPLICATION NUMBER: US/09/408,020  
CURRENT FILING DATE: 1999-09-29  
PRIOR APPLICATION NUMBER: 60/102,294  
PRIOR FILING DATE: 1998-09-29  
NUMBER OF SEQ ID NOS: 123  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 68  
LENGTH: 92  
TYPE: PRT  
ORGANISM: Cenarchaeum symbiosum  
US-09-408-020-68

Query Match 21.1%; Score 179.5; DB 4; Length 92;  
Best Local Similarity 40.4%; Pred. No. 4.4e-12;  
Matches 44; Conservative 19; Mismatches 23; Indels 23; Gaps 4;

QY 54 RVRCSHLVXHSQRRSSWFQEKITRKEALELINGYQIKSGEDFESLASQFS-D 112  
DB 4 KIKCSHLVX-----KQGEALAV----QERLKAGEK-FGKLAKELSID 41

QY 113 CSSAKARGDLGAFSRGCMQKPEPDASFALRTGEMSGPVFTDSGIHLIR 161  
DB 42 GGSAKRDSGLGYFGRGMVKPFPEDAAAFRLQVGEVSEPVKSEFGYHVIKR 90  
RESULT 12  
US-09-252-404A-1  
Sequence 1, Application US/09252404A  
Patent No. 6495376  
GENERAL INFORMATION:  
APPLICANT: Kun Ping Lu  
APPLICANT: Xiao Zhen Zhou  
TITLE OF INVENTION: Methods and Compositions for Regulating Protein-Protein Interactions  
FILE REFERENCE: 1440.1034-000  
CURRENT APPLICATION NUMBER: US/09/252,404A  
CURRENT FILING DATE: 1999-02-18  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 31  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-252-404A-1

Query Match 20.2%; Score 172; DB 4; Length 31;  
Best Local Similarity 96.8%; Pred. No. 6.5e-12;  
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 LPFGWEKMSRSGRVVYFNHITNASQWERP 37  
DB 1 LPFGWEKMSRSGRVVYFNHITNASQWERP 31

RESULT 13  
US-09-208-804-1  
Sequence 1, Application US/09208804  
Patent No. 6030826  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
TITLE OF INVENTION: NOVEL HUMAN PARVULIN-LIKE PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/208,804  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/801,743  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0217 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 131 amino acids  
TYPE: amino acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
US-09-208-804-1

Query Match          17.8%; Score 152; DB 3; Length 131;
Best Local Similarity 34.0%; Pred. No. 6.8e-09;
Matches 51; Conservative 14; Mismatches 33; Indels 52; Gaps 9;

QY 37 PSNSSSG-----GKNGGEPARVRCSHLLV-KHSQSRPSSMRQE 76
Db 3 PKGSGSGKAGKGAASGSDSADKKAQPKGGNAVKR--HILCEKHG----- 49

QY 77 KITRTKEALELINGYIQIKSGEEDFESLASQFDCSSAKARGDLGAFSRGOMQKPPED 136
Db 50 KIM-----EAME-----KLKSGMR-FNEVAAQYSE-DKARQGGDLGWNTRGSMVGPFOE 96

QY 137 ASFALRTGEMSGPVFTDS-----GIHIL 160
Db 97 AAFALPVSGMDKPVFTDPPVTKFGYHIIM 126

RESULT 15
US-08-801-743-1
; Sequence 1, Application US/08801743
; Patent No. 6037164
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: NOVEL HUMAN PARVULIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,743
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0217 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
US-08-801-743-1

Query Match          17.8%; Score 152; DB 3; Length 131;
Best Local Similarity 34.0%; Pred. No. 6.8e-09;
Matches 51; Conservative 14; Mismatches 33; Indels 52; Gaps 9;
```

```
QY 37 PSNSSSG-----GKNGGEPARVRCSHLLV-KHSQSRPSSMRQE 76
Db 3 PKGSGSGKAGKGAASGSDSADKKAQPKGGNAVKR--HILCEKHG----- 49

QY 77 KITRTKEALELINGYIQIKSGEEDFESLASQFDCSSAKARGDLGAFSRGOMQKPPED 136
Db 50 KIM-----EAME-----KLKSGMR-FNEVAAQYSE-DKARQGGDLGWNTRGSMVGPFOE 96

QY 137 ASFALRTGEMSGPVFTDS-----GIHIL 160
Db 97 AAFALPVSGMDKPVFTDPPVTKFGYHIIM 126

RESULT 15
US-09-621-976-7038
; Sequence 7038, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Joibert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTS and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 7038
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-7038

Query Match          17.8%; Score 152; DB 4; Length 131;
Best Local Similarity 34.0%; Pred. No. 6.8e-09;
Matches 51; Conservative 14; Mismatches 33; Indels 52; Gaps 9;

QY 37 PSNSSSG-----GKNGGEPARVRCSHLLV-KHSQSRPSSMRQE 76
Db 3 PKGSGSGKAGKGAASGSDSADKKAQPKGGNAVKR--HILCEKHG----- 49

QY 77 KITRTKEALELINGYIQIKSGEEDFESLASQFDCSSAKARGDLGAFSRGOMQKPPED 136
Db 50 KIM-----EAME-----KLKSGMR-FNEVAAQYSE-DKARQGGDLGWNTRGSMVGPFOE 96

QY 137 ASFALRTGEMSGPVFTDS-----GIHIL 160
Db 97 AAFALPVSGMDKPVFTDPPVTKFGYHIIM 126
```

Search completed: June 28, 2004, 18:40:37  
Job time : 31 secs

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 5.1.1.6

Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:24 ; Search time 2080.37 Seconds  
(without alignments)  
9875.458 Million cell updates/sec

Title: AAC50492  
Perfect score: 474  
Sequence: 1 ATGGCTGCTACGAGCAAC.....TTCTGGAGCGCTGCTTAG 474

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Result No.	Score	Match	Length	DB	ID	Description
1	301.2	63.5	621	8	AY086702	Arabidops
2	233	43.2	411	8	AY097375	Arabidops
3	233	49.2	594	8	AY070749	Arabidops
4	226.2	47.7	617	8	MSU09465	Medicago sa
5	224.6	47.4	660	8	MSU09464	Medicago sa
6	224.6	47.4	679	8	MSU09461	Medicago sa
7	224.6	47.4	693	8	MSU09460	Medicago sa
8	223.4	47.1	102183	8	AC027134	Arabidops
9	223.4	47.1	103157	8	AC011810	Arabidops
10	220.4	46.5	96379	8	ATT5K6	Arabidops
11	220.4	46.5	98473	8	ATT3OR20	Arabidops
12	218.2	46.0	125590	3	CEY49B10	Arabidops
13	214.8	45.3	4534	3	DHHS334	Caenorhabdi
14	210	44.3	119378	2	AC130200	Medicago
15	208.6	44.0	411	8	AY078027	Arabidops
16	208.6	44.0	706	8	AF385735	Arabidops
17	206	43.5	129547	8	AC006434	Genomic s
18	205.6	43.4	571	8	MSU09462	Medicago sa
19	203	42.8	595	8	MSU09463	Medicago sa
20	203	42.8	723	8	AB015760	Nicotiana
21	202.2	42.7	922	8	MSU09459	Medicago sa
22	200.8	42.4	699	6	AX364448	Sequence
23	200.4	42.3	113431	8	AC119413	Medicago
24	199.8	42.2	132883	2	AC146875	Brassica
25	199	42.0	717	8	AY086668	Arabidops
26	199	42.0	729	6	AX364455	Sequence
27	199	42.0	792	8	BT002391	Arabidops
28	199	42.0	834	8	BT003326	Arabidops
29	198.8	41.9	114331	8	AC119413	Medicago
30	197.4	41.6	826	8	AK120807	Oryza sat
31	197.4	41.6	2339	8	AK100960	Oryza sat
32	195.2	41.2	585	6	AX364443	Sequence
33	194.8	41.1	579	8	MSHS3C3	Medicago sa
34	194.6	41.1	411	8	AY077654	Arabidops
35	194.6	41.1	655	8	AY037250	Arabidops
36	194.6	41.1	70952	8	AB024028	Arabidops
37	194.2	41.0	618	8	MSHS3C1	Medicago sa
38	194.2	41.0	668	8	LEH33	Medicago sa
39	194.2	41.0	1089	8	MSHS311	Arabidops
40	193.4	40.8	411	8	AF370577	Arabidops
41	193.4	40.8	471	8	AF370577	Arabidops
42	193.4	40.8	626	8	AY039904	Arabidops
43	193.4	40.8	669	6	AX364449	Sequence
44	193.4	40.8	682	6	AX364072	Sequence
45	193.4	40.8	682	6	AX366162	Sequence

ALIGNMENTS

RESULT 1  
AY086702  
LOCUS Arabidopsis thaliana clone 268982 mRNA, complete sequence.  
DEFINITION Arabidopsis thaliana clone 268982 mRNA, complete sequence.  
ACCESSION AY086702  
VERSION AY086702.1 GI:21405412  
KEYWORDS FLI CDNA..  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (bases 1 to 621)  
Haas, B.J., Volkovskiy, N., Town, C.D., Troukhan, M., Alexandrov, N.,  
Feidmann, K.A., Flavell, R.B., White, O. and Salzberg, S.L.

Pred. No. is the number of results predicted by chance to have a



TITLE Full-length messenger RNA sequences greatly improve genome annotation  
JOURNAL Genome Biol. 3 (6), RESEARCH0029 (2002)  
MEDLINE 22088475  
PUBMED 12093376  
REFERENCE 2 (bases 1 to 621)  
AUTHORS Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.  
TITLE Full-length cDNA from Arabidopsis thaliana  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 621)  
AUTHORS Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.  
TITLE Direct Submission  
JOURNAL Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA  
COMMENT This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or Laer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genet carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.

FEATURES  
source  
1..621  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone="268982"  
92..502  
/codon\_start=1  
/product="histone H3 protein, putative"  
/protein\_id="AA063756.1"  
/db\_xref="GI:21555021"  
/translation="MARTQTARKSFGKAPTKQLATKARKSAPTTGGVKKPHERP  
GVVALREIRYQKSTELLNRKLPFQRLVREIAQDFKIDLFQSHAVLQEAAYLV  
GLFEDNLCAIHAQRVTIMPQVQLARRAERA"

ORIGIN  
Query Match 63.5%; Score 301.2; DB 8; Length 621;  
Best Local Similarity 81.9%; Pred. No. 5.6e-85;  
Matches 389; Conservative 0; Mismatches 23; Indels 63; Gaps 1;  
1 ATGGCTCGTACGAGCAAACTGCAAGAAAATCACCGGAGAAAAGTCCGAGGACTCTG 60  
2 ATGGCTCGTACGAGCAAACTGCAAGAAAATCACCGGAGAAAAGTCCGAGGAGCAG 151  
61 CTGCTACCAAGGGCGGAGAAATCGCGCGACTACTCGAGAGTCAGAACTCCAC 120  
152 CTGCTACCAAGGGCGGAGAAATCGCGCGACTACTCGAGAGTCAGAACTCCAC 211  
121 CGTTTACCGTCCCGAAACCGTCGCTCTTCTGAGATTCGTAATACCAAGAGACACAG 180  
212 CGTTTCCGTCCTGGAACCGTCGCTCTTCTGAGATTCGTAATACCAAGAGACACAG 271  
181 TTGTTGATCGGTAAACTTCCTTTTCAACGCTGCTTCTGTAAPCGCTCAAGATTACAAG 240  
272 TTGTTGACCGTAAACTTCATTCCAAAGCTTCTGTTGTAATCGCTCAAGATTCAA- 330  
241 ACGGATCTGAGATTCGAGGCCATCGGCTGTAGCTCTTCAAGAGCTGCTGAGCATAT 300  
331 -----  
301 TTGACGATCTGAGATTCGAGGCCATCGGCTGTAGCTCTTCAAGAGCTGCGGAGCA 360

Db 331 --GACGGATCTGAGATTCCAAAGCCATCGAGTGTAGTCTTCAAGAGCTCGGAGCA 388  
QY 361 TATTTGGTGGGTTTGTGTTGAAGACACAAATCTGTGCCATTTCATGCAAGAGGTTACG 420  
Db 389 TATTTGGTGGTGTGTTGTTGAAGACACAAATCTGTGCCATTTCATGCAAGAGGTTACC 448  
QY 421 ATATGCTTAAGATGTTCAATTCGCAAGAGGATTCGTGAGAGCGCTGCTTAG 474  
Db 449 ATTATGCTTAAGATGTTCAATTCGCAAGAGGATTCGTGAGAGCGCTGCTTAG 502

RESULT 2  
LOCUS AY097375 411 bp mRNA linear PLN 05-MAY-2002  
DEFINITION Arabidopsis thaliana AT5g10980/T30N20\_250 mRNA, complete cds.  
ACCESSION AY097375  
VERSION AY097375.1 GI:20453304  
KEYWORDS FLI CDNA.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
REFERENCE 1 (bases 1 to 411)  
AUTHORS Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu.S., Shinozaki,K., Davis,R.W., Theologis,A., and Ecker,J.R.  
TITLE Arabidopsis ORF clones  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 411)  
AUTHORS Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu.S., Shinozaki,K., Davis,R.W., Theologis,A., and Ecker,J.R.  
TITLE Direct Submission  
JOURNAL Submitted (15-APR-2002) Salk Institute Genomic Analysis Laboratory (SIGNAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.  
The Salk, Stanford, PGEN (SSP) Consortium members constructed and sequenced the pUNI (PGEN) clones using the RAFL cDNAs: Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L., Chang,E., Dale,J.M., Goldsmith,A.D., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu.S., Davis,R.W., Theologis,A., and Ecker,J.R.  
Shinn,P. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk) contributed equally to this work as PIs.

FEATURES  
source  
1..411  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="5"

```

/clone="U01392"
/ecotype="Columbia"
/note="this clone is in pUNI 51"
1. .411
/note="histone H3 protein"
/codon_start=1
/product="AT5g10980/T3ON20_250"
/protein_id="AAMI19891.1"
/db_xref="GI:20453305"
/translation="MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRP
GTVALRIRKYQKSTELLIRKLPFQRLVREIAQDFKIDLRFSQSHAVLALQEAABAYLV
GLPFDTNLCALHAKXRVIMPDKIQLARRIGERA"

ORIGIN
Query Match 49.2%; Score 233; DB 8; Length 411;
Best Local Similarity 72.9%; Pred. No. 3.8e-63;
Matches 345; Conservative 0; Mismatches 65; Indels 63; Gaps 1;

Qy 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATACACAGGAGGAAAGCTCCGAGGACTCTG 60
Db |||||
Qy 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATACACAGGAGGAAAGCTCCGAGGAGGAG 60
Db |||||
Qy 61 CTCGCTACCAAGGGCGGAGGAAATCTGCCGCACTACTGAGGAGTCAAGAAACCTCAC 120
Db |||||
Qy 61 CTTGCCACCAAGGGCGGAGGAAATCAGCAACCAACCGGAGGAGTCAAGAGCCTCAC 120
Db |||||
Qy 121 CGTTAGCGTCCGGAACCGTCTGCTTCTGAGAGTTCGTAATACCAGAGGACAGAG 180
Db |||||
Qy 121 CGTTAGCGTCCGGAACCGTCTGCTTCTGAGAGTTCGTAATACCAGAGGAGTACTGAA 180
Db |||||
Qy 181 TTGTTGATCCGTAACCTCTTTCAAGCTCTGTTCTGTAATCGCTCAAGATTACAAAG 240
Db |||||
Qy 181 TTGTTGATCCGTAACCTCTTTCAAGCTCTGTTCTGTAATCGCTCAAGATTACAAAG 240
Db |||||
Qy 241 ACGGATCTGAGATTCAGAGGAGTCCGTTGTTAGCTCTTCAAGAGCTGCTCAAGCAT 300
Db |||||
Qy 241 ACGGATCTGAGATTCAGAGTCAAGGATCGGTTGTTAGCTCTTCAAGAGGCGCT 291
Db |||||
Qy 301 TTGAGGATCTGAGATTCAGAGGATCGGTTGTTAGCTCTTCAAGAGGCTGGGAGCA 360
Db |||||
Qy 292 -----GAGGCA 297
Db |||||
Qy 361 TATTTGGTGGGTTGTTGAAGACAAATCTGTGCAATTCATGCAAGAGGCTTACG 420
Db |||||
Qy 298 TACCTAGTTGACTTTTCGAGATCTACTCTCTGCTATTCAGCCCAAGAGGCTTACG 357
Db |||||
Qy 421 ATAATGCTTAAGATGTTCAATGTCAGAGGATTCGTTGAGAGCGGTCTTA 473
Db |||||
Qy 358 ATCATGCTTAAGACATTCATTTGGCTAGAGAAATTCGTTGAGAGCGTGTCTTA 410
Db |||||

RESULT 3
AY070749 594 bp mRNA linear PLN 26-DEC-2001
LOCUS Arabidopsis thaliana AT5g10980/T3ON20_250 mRNA, complete cds.
DEFINITION Arabidopsis thaliana (thale cress)
ACCESSION AY070749
VERSION AY070749.1 GI:17979504
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 594)
AUTHORS Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayaishizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.

Arabisidopsis cDNA clones

```

## JOURNAL REFERENCE

Unpublished  
2 (bases 1 to 594)  
Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayaishizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.

## TITLE JOURNAL

Direct Submission  
Submitted (19-DEC-2001) Salk Institute Genomic Analysis Laboratory (SIGNAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

## COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA': Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayaishizaki,Y. and Shinozaki,K.).

The Salk, Stanford, PBC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Shinn,P., Banh,J., Bowser,L., Dale,J.M., Goldsmith,A.D., Jiang,P.X., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K., Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Davis,R.W., Theologis,A., and Ecker,J.R.

Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk) contributed equally to this work as PIs.

## FEATURES

Location/Qualifiers

1. .594  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="RAFL06-13-E17(R21392)"  
/ecotype="Columbia"  
1. .92

## 5'UTR

CDS  
93. .503  
/note="histone H3 protein"  
/codon\_start=1  
/product="AT5g10980/T3ON20\_250"  
/protein\_id="AALS0086.1"  
/db\_xref="GI:17979505"  
/translation="MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRP  
GTVALRIRKYQKSTELLIRKLPFQRLVREIAQDFKIDLRFSQSHAVLALQEAABAYLV  
GLEPDTNLCALHAKXRVIMPDKIQLARRIGERA"  
504. .594

## 3'UTR

## ORIGIN

Query Match 49.2%; Score 233; DB 8; Length 594;  
Best Local Similarity 72.9%; Pred. No. 4.1e-63;  
Matches 345; Conservative 0; Mismatches 65; Indels 63; Gaps 1;  
Qy 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATACACAGGAGGAAAGCTCCGAGGACTCTG 60  
Db |||||  
Qy 93 ATGGCTCGTACGAAGCAAACTGCAAGAAATACACAGGAGGAAAGCTCCGAGGAGGAG 152  
Db |||||  
Qy 61 CTGCTACCAAGGGCGGAGGAAATCTGCCGCACTACTGAGGAGTCAAGAAACCTCAC 120  
Db |||||  
Qy 153 CTGTCACCAAGGGCGGAGGAAATCAGACCAACCGGAGGAGTCAAGAGCTCAC 212  
Db |||||  
Qy 121 CGTTACCGTCCGGAACCGTCTGCTTCTGAGAGTTCGTAATACCAGAGGACAGAG 180  
Db |||||  
Qy 213 CGTTACCGTCCGGAACCGTCTGCTTCTGAGAGTTCGTAATACCAGAGGAGTACTGAA 272  
Db |||||  
Qy 181 TTGTTGATCCGTAACCTCTTTCAAGCTCTGTTCTGTAATCGCTCAAGATTACAAAG 240  
Db |||||  
Qy 273 TTGTTGATCCGTAACCTCTTTCAAGCTCTGTTCTGTAATCGCTCAAGATTACAAAG 332  
Db |||||

## TITLE

```

241 ACGGATCTGAGATTCAGAGCCATCGGTTGTTAGTCTTTCAGAAAGCTGCTGAAGCATAT 300
333 ACGGATCTGAGGTTTCAGAGTCAAGGCTGTTAGTCTTTCAGAAAGCGGCT----- 383
301 TTGACGGATCTGAGATTCAGAGCCATCGGTTGTTAGTCTTTCAGAAAGCTGCTGAAGCATAT 360
384 -----GAGGCA 389
361 TATTTGGTGGTTTGTTCAGAGACAAATCTGTGTCATTCATCAAGAGGGTTACG 420
390 TACCTAGTTGGACTTTTCAGGATCTATCTCTGCGCTATTACGCCAAGAGGGTTACG 449
421 ATAATGCCTAAGATGTTCAATGTCGCAAGAGATTCGTGGAGAGCGTCTTAA 473
450 ATCATGCTTAAGACATTCATTAATTTGGCTAGAGAAATTCGTGGAGAACGTGCTTAA 502

RESULT 4
MSU09465 617 bp mRNA linear PLN 23-NOV-1996
LOCUS Medicago sativa Regen S clone p33cl31 histone H3.2 mRNA, complete
DEFINITION cds.
ACCESSION U09465
VERSION U09465.1 GI:488576
KEYWORDS Medicago sativa
SOURCE Medicago sativa
ORGANISM Medicago sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 617)
AUTHORS Robertson,A.J.
TITLE Histone H3 genes in alfalfa
JOURNAL Thesis (1994) Division of Cell Biology and Biophysics, University
of Missouri-Kansas City
2 (bases 1 to 617)
AUTHORS Robertson,A.J., Kepros,T., Dudits,D. and Waterborg,J.H.
TITLE Identification of three highly expressed replacement histone H3
genes of alfalfa
JOURNAL DNA Seq. 6 (3), 137-146 (1996)
MEDLINE 96293858
PUBMED 8722568
REFERENCE 3 (bases 1 to 617)
AUTHORS Robertson,A.J.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-1994) Anthony J. Robertson, University of
Missouri-Kansas City, Division of Cell Biology and Biophysics, 5100
Rockhill Road, Kansas City, MO 64110, USA
FEATURES
source
1. .617
/organism="Medicago sativa"
/mol_type="mRNA"
/strain="Regen S"
/db_xref="taxon:3879"
/clone="p33cl31"
/dev_stage="embryo"
<1. .17
/evidence=experimental
18. .428
/codon_start=1
/evidence=experimental
/product="histone H3.2"
/protein_id="AAB36498.1"
/db_xref="GI:488577"
/translaton="MARTKOTARKSTGKAPRKOLATKAARKSAPTTGGVKPHRYRP
GTVALREIKYQKSTELLIRKLPORLVREIAODFKTLRFQSHAVLALQEAAEAVLV
GLPFDNLCAIHAKRVIMPDIQLARRIGERA"
21. .425
/product="unnamed"
/evidence=experimental

variation
53 /note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replace="t"
242
variation
242 /note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replace="c"
264
variation
264 /note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replace="c"
429. .613
3'UTR /evidence=experimental
449
variation
449 /note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replace="c"
527. .576
repeat_region /evidence=experimental
597. .603
polya_signal /rpt_unit="527. .551"
614 /evidence=experimental
polya_site /evidence=experimental

ORIGIN
Query Match 47.7%; Score 226.2; DB 8; Length 617;
Best Local Similarity 84.2%; Pred. No. 6.2e-61;
Matches 255; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
QY 1 ATGGCTCTGATACGAAGCAAACTCCAGAGAAATCACAGGAGGAAAAGTCCGAGGACTCTG 60
DB 18 ATGGCTCTGATACGAAGCAAACTCCAGAGAAATCACAGGAGGAAAAGTCCGAGGACTCTG 77
QY 61 CTCGGTCTACCAAGCGCGGAGGAAATCTGCGCGACTCTCTGGAGGAGTCAAGAAACCTCAC 120
DB 78 CTCGCCACCAAGGCTGCTAGGAAATCTGCTCTCTACTCTGGAGGAGTCAAGAAACCTCAC 137
QY 121 CGTTACCGTCCCGGAACCGTCTGCTCTCTGAGATTCGTAATACCAAGAGACAGAG 180
DB 138 CGATACCGGCTCGAACTGCTGCTCTCTGAGATTCGTAATACCAAGAGACAGAG 197
QY 181 TTGTTGATCGGTAAGTCTCTTTTCAAGCTCTTTTGGTGAATCGCTCAAGATTACAAG 240
DB 198 CTTTGTATCGCAAGCTTCCATTTAGCGTCTTTCGCGTGAATCGCTCAAGATTACAAG 257
QY 241 ACGGATCTGAGATTCAGAGCCATCGGTTGTTAGTCTTTCAGAAAGCTGCTGAAGCATAT 300
DB 258 ACGGATTCAGATTCAGAGCCATCGGTTGTTAGTCTTTCAGAAAGCTGCTGAAGCATAT 317
QY 301 TTG 303
DB 318 CTG 320

RESULT 5
MSU09464 660 bp mRNA linear PLN 23-NOV-1996
LOCUS Medicago sativa Regen S clone p33cl30 histone H3.2 mRNA, complete
DEFINITION cds.
ACCESSION U09464
VERSION U09464.1 GI:488574
KEYWORDS Medicago sativa
SOURCE Medicago sativa
ORGANISM Medicago sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 660)
AUTHORS Robertson,A.J.
mat_peptide

```

TITLE  
 JOURNAL  
 Histone H3 genes in alfalfa  
 Thesis (1994) Division of Cell Biology and Biophysics, University  
 of Missouri-Kansas City  
 2 (bases 1 to 660)  
 REFERENCE  
 AUTHORS  
 Roberton,A.J., Kapros,T., Dudits,D. and Waterborg,J.H.  
 Identification of three highly expressed replacement histone H3  
 genes of alfalfa  
 TITLE  
 JOURNAL  
 DNA Seq. 6 (3), 137-146 (1996)  
 MEDLINE  
 PUBMED  
 96293858  
 REFERENCE  
 3 (bases 1 to 660)  
 AUTHORS  
 Roberton,A.J.  
 Direct Submission  
 TITLE  
 JOURNAL  
 Submitted (06-MAY-1994) Anthony J. Robertson, University of  
 Missouri-Kansas City, Division of Cell Biology and Biophysics, 5100  
 Rockhill Road, Kansas City, MO 64110, USA  
 FEATURES  
 source  
 1. .660  
 /organism="Medicago sativa"  
 /mol\_type="mRNA"  
 /strain="Regen S"  
 /db\_xref="taxon:3879"  
 /clone="pH3cl130"  
 /clone\_lib="LUCA"  
 /dev\_stage="embryo"  
 <1. .17  
 /evidence=experimental  
 18. .428  
 /codon\_start=1  
 /evidence=experimental  
 /product="histone H3.2"  
 /protein\_id="AAB36497.1"  
 /db\_xref="GI:488575"  
 /translation="MARTQTAKSTGSKAPRKQLATKAARKSAPTTGGVKKPHRYRP  
 GVVALREIKYQKSTELLRLKLPFORLVRLEIAODFKTLRFQSHAVLALQEAAYLV  
 GTFEDNLCAIHKRVTIWPKIQLARRIGERA"  
 21. .425  
 /product="unnamed"  
 /evidence=experimental  
 155  
 /notes="Sequence variation with genomic H3.2 clone msh3g1,  
 GenBank Accession Number U09458"  
 /replace="t"  
 264  
 /notes="Sequence variation with genomic H3.2 clone msh3g1,  
 GenBank Accession Number U09458"  
 /replace="c"  
 429. .655  
 /evidence=experimental  
 449  
 /notes="Sequence variation with genomic H3.2 clone msh3g1,  
 GenBank Accession Number U09458"  
 /replace="c"  
 499  
 /notes="Sequence variation with genomic H3.2 clone msh3g1,  
 GenBank Accession Number U09458"  
 /replace="t"  
 584. .607  
 /evidence=experimental  
 /rpt\_type=DIRECT  
 /rpt\_unit="584. .595"  
 609. .615  
 611. .628  
 628. .634  
 /evidence=experimental  
 655  
 /evidence=experimental  
 5'UTR  
 CDS  
 mat\_peptide  
 variation  
 variation  
 3'UTR  
 variation  
 variation  
 repeat\_region  
 polyA\_signal  
 stem\_loop  
 polyA\_signal  
 polyA\_site  
 ORIGIN  
 Query Match 47.4%; Score 224.6; DB 8; Length 660;  
 Best Local Similarity 83.8%; Pred. No. 2e-60;  
 Matches 254; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 ATGGCTCTACGAGCAAACTGCAGAAAATCACACGGGAGGAAAGCTCCGAGGACTCTG 60  
 Db 18 ATGGCTCTACGAGCAAACTGCAGAAAATCACACGGGAGGAAAGCTCCGAGGAGCAG 77  
 QY 61 CTCGCTACCAAGGGCGGAGGAAATCTGCGCGGACTACTTGGAGGAGTCAAGAAACCTCAC 120  
 Db 78 CTCGCGACCAAGGCTGCTAGGAAATCTGCTCTACTTCTGAGGAGTCAAGAAACCTCAC 137  
 QY 121 CGTTACCGTCCCGAAACCGTCTCTTCTGAGATTCGTAATACAGAAAGCACAGAG 180  
 Db 138 CGATACCGCCCTGGAACCGTCTCTTCTGAGATTCGTAATACAGAAAGCACAGAG 197  
 QY 181 TTGTTGATCCGTAACCTCTTTTCAACGTCCTTGTGCTGTAATCGCTCAAGATTACAAG 240  
 Db 198 CTTTGTATCCGCAAGCTTCCATTTTCAGGCTCTTGTCCGTTGAAATTTGCTCAAGATTCAAG 257  
 QY 241 ACGGATCTGAGATTCAGAGCCATCGGCTGTTAGTCTTCTCAAGAAAGCTGCTGAAGCATAT 300  
 Db 258 ACGGATTGAGATTCAGAGCCATCGAGTCTTTCACATTCAGAAAGCACGCTGAGGCTTAC 317  
 QY 301 TTG 303  
 Db 318 CTG 320  
 RESULT 6  
 MSU09461  
 LOCUS  
 DEFINITION  
 Medicago sativa Regen S clone pH3cl18 histone H3.2 mRNA, complete  
 cds.  
 ACCESSION  
 U09461  
 VERSION  
 U09461.1  
 GI:488568  
 KEYWORDS  
 Medicago sativa  
 ORGANISM  
 Medicago sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.  
 REFERENCE  
 1 (bases 1 to 679)  
 AUTHORS  
 Roberton,A.J.  
 TITLE  
 Histone H3 genes in alfalfa  
 JOURNAL  
 Thesis (1994) Division of Cell Biology and Biophysics, University  
 of Missouri-Kansas City  
 2 (bases 1 to 679)  
 REFERENCE  
 AUTHORS  
 Roberton,A.J., Kapros,T., Dudits,D. and Waterborg,J.H.  
 TITLE  
 Identification of three highly expressed replacement histone H3  
 genes of alfalfa  
 JOURNAL  
 DNA Seq. 6 (3), 137-146 (1996)  
 MEDLINE  
 PUBMED  
 96293858  
 REFERENCE  
 3 (bases 1 to 679)  
 AUTHORS  
 Roberton,A.J.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (06-MAY-1994) Anthony J. Robertson, University of  
 Missouri-Kansas City, Division of Cell Biology and Biophysics, 5100  
 Rockhill Road, Kansas City, MO 64110, USA  
 FEATURES  
 source  
 1. .679  
 /organism="Medicago sativa"  
 /mol\_type="mRNA"  
 /strain="Regen S"  
 /db\_xref="taxon:3879"  
 /clone="pH3cl18"  
 /clone\_lib="LUCA"  
 /dev\_stage="embryo"  
 <1. .41  
 /notes="A complete 5'UTR is present in pH3cl10, GenBank  
 Accession Number U09460"  
 /evidence=experimental  
 1. .34  
 /evidence=experimental  
 /rpt\_family="CT rich"  
 5'UTR  
 repeat\_region

```

/rpt_type=DIRECT
/rpt_unit="1. .17"
42. .452
/codon_start=1
/evidence=experimental
/product="histone H3.2"
/protein_id="AAB36494.1"
/db_xref="GI:488569"
/translaton="MARTKOTARKSTGKAPRKOLATKAARKSAPTTGGVKKPHRYRP
GTVALRIRKYQKSTELLIRKLPQRLVREIAQDFKTDLRFSQSHAVLALQBAEAAYLV
GLFEDTNLCATHAKRVITMPKDIQLARRINGERA"
45. .449
/product="unnamed"
/evidence=experimental
179
/note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replacement="c"
288
/note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replacement="c"
453. .679
/evidence=experimental
473
/note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replacement="c"
523
/note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replacement="c"
608. .631
/evidence=experimental
/rpt_type=DIRECT
/rpt_unit="608. .619"
633. .638
635. .652
/note="The stem-loop is located among the two
polyadenylation signals"
652. .658
/evidence=experimental
679
/note="20 A residues"
/evidence=experimental

ORIGIN
Query Match 47.4%; Score 224.6; DB 8; Length 679;
Best Local Similarity 83.8%; Pred. No. 2e-60;
Matches 254; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 ATGGCTGTACGAAGCAAACTCGAGAAATACACGGAGGAAAGCTCCGAGGACTCTG 60
Db |||||
42 ATGGCTGTACGAAGCAAACTCGAGAAATACACGGAGGAAAGCTCCGAGGAGCAG 101
QY 61 CTCGCTACCAAGCGCGAGGAAATCTCGCGGACTACTGGAGGAGTCAAGAAACCTCAC 120
Db |||||
102 CTCGCCACCAAGGCTGCTAGGAAATCTGCTCTACTACTGGAGGAGTCAAGAAACCTCAC 161
QY 121 CGTTACCGTCCCGGAACCGTCTCTCTGAGATTTCGTAATACAGAGAGCAGCAGAG 180
Db |||||
162 CGATACCGCCTCGAACCGTCTCTCTGAGATTCGTAATACAGAGAGTACCGAG 221
QY 181 TTGTTGATCGGTAACCTCTCTTCAAGCTCTGTTGCTGGAATCGCTCAAGATTACAAG 240
Db |||||
222 CTTTTGTATCGCAAGCTTCATTTTCAGCGCTTGTTCGCTGGAATTCCTCAAGATTTCAAG 281
QY 241 ACGGATCTGAGATTCCAGAGCCATCGGCTGTTAGCTCTTCAAGAAAGCTCTGAAGCATAT 300
Db |||||
282 ACGGATTGAGATTCAGAGCCATGAGTCTTGGACCTTCAGAGAGCAGTGAAGCTTAC 341
QY 301 TTG 303
||

```

```

Db 342 CTG 344

RESULT 7
MSU09460 693 bp mRNA linear PLN 23-NOV-1996
LOCUS Medicago sativa Regen S clone ph3c110 histone H3.2 mRNA, complete
DEFINITION
ACCESSION U09460
VERSION U09460.1 GI:488566
KEYWORDS
SOURCE Medicago sativa
ORGANISM Medicago sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 693)
AUTHORS Roberton,A.J.
TITLE Histone H3 genes in alfalfa
JOURNAL Thesis (1994) Division of Cell Biology and Biophysics, University
of Missouri-Kansas City
REFERENCE 2 (bases 1 to 693)
AUTHORS Roberton,A.J., Kaptos,T., Dudits,D. and Waterborg,J.H.
TITLE Identification of three highly expressed replacement histone H3
genes of alfalfa
JOURNAL DNA Seq. 6 (3), 137-146 (1996)
MEDLINE 96293858
PUBMED 8722568
REFERENCE 3 (bases 1 to 693)
AUTHORS Roberton,A.J.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-1994) Anthony J. Roberton, University of
Missouri-Kansas City, Division of Cell Biology and Biophysics, 5100
Rockhill Road, Kansas City, MO 64110, USA

FEATURES
source
1. .693
/organism="Medicago sativa"
/mol_type="mRNA"
/strain="Regen S"
/db_xref="taxon:3879"
/clone="ph3c110"
/clone_lib="LUCa"
/dev_stage="embryo"
1. .55
/evidence=experimental

5'UTR
repeat_region
15. .48
/note="Repeat is present as a single sequence element in
the 5' UTR of the known genomic clone msH3g1, GenBank
Accession Number U09458"
/rpt_family="CT-rich"
/rpt_type=DIRECT
/rpt_unit="15. .31"
56. .466
/codon_start=1
/evidence=experimental
/product="histone H3.2"
/protein_id="AAB36493.1"
/db_xref="GI:488567"
/translaton="MARTKOTARKSTGKAPRKOLATKAARKSAPTTGGVKKPHRYRP
GTVALRIRKYQKSTELLIRKLPQRLVREIAQDFKTDLRFSQSHAVLALQBAEAAYLV
GLFEDTNLCATHAKRVITMPKDIQLARRINGERA"
59. .463
/product="unnamed"
/evidence=experimental
193
/note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replacement="c"
302
/note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replacement="c"

```

mat\_peptide  
variation  
variation

```

3'UTR          457..693
/sequence=experimental
variation      487
GenBank Accession Number U09458
/replacement="c"
variation      537
/sequence=variation with genomic H3.2 clone msh3g1,
GenBank Accession Number U09458
/replacement="t"
repeat_region 622..645
/sequence=Repeat is present as a single sequence element in
the 3'UTR of the known gene sequence of msh3g1, GenBank
Accession Number U09458
/rpt_type=DIRECT
/rpt_unit="622..633"
polyA_signal  647..653
stem_loop     649..666
/sequence="The stem-loop is located among the two
polyadenylation signals"
polyA_signal  666..672
/sequence=experimental
polyA_site    693
/sequence="20 A residues"
/sequence=experimental

```

## ORIGIN

```

Query Match      47.4%; Score 224.6; DB 8; Length 693;
Best local Similarity 83.8%; Pred. No. 2.1e-60;
Matches 254; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 ATGGCTCTGACGACGAACTGCAAGAAATTCACACGAGGAGAAAGCTCCGAGACTCTG 60
   |||||
DB 56 ATGGCTGTGTACAGCAACCGCTGCGAAATCCACTGGTGTAGGCTCCAGGAGCAG 115
   |||||
QY 61 CTCGCTACCAAGGGCGGAGAAATTCGCGCGACTACTGAGGAGTCAAGAAACCTCAC 120
   |||||
DB 116 CTCGCCACCAAGGCTGTAGGAAATCTCTACTACTGAGGAGTCAAGAAACCTCAC 175
   |||||
QY 121 CGTTACGCTCCGGAACCGTCTGCTGAGATTCGTAATACCAAGAGCAGCAGAG 180
   |||||
DB 176 CGATACGCGCTGGAACCGTCTGCTGAGATTCGTAATACCAAGAGCAGCAGAG 235
   |||||
QY 181 TTGTTGATCCCTAACTTCCTTTTCAAGCTCTGTTCTGTAATCGCTCAAGATTACAAG 240
   |||||
DB 236 CTTTGTATCCGCAAGCTTCCATTCAGCGCTCTGTGCGTAATGCTCAAGATTCAAG 295
   |||||
QY 241 ACGGATCTGAGATTCAGAGCCATCGGCTGTAGCTTTCAGAAAGCTGCTGAAGCATAT 300
   |||||
DB 296 ACGGATTTGAGATTCAGAGCCATCGGCTGTAGCTTTCAGAAAGCTGCTGAAGCATAT 355
   |||||
QY 301 TTG 303
   |||
DB 356 CTG 358

```

```

RESULT 8
AC027134/c
LOCUS          102183 bp      DNA      linear      PLN 15-AUG-2000
DEFINITION    Arabidopsis thaliana chromosome I BAC F13B4 genomic sequence,
complete sequence.
ACCESSION     AC027134
KEYWORDS      HTG
SOURCE        Arabidopsis thaliana (thale cress)
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 102183)
FEDERSPIEL,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
ALTAFI,H., Nguyen,M., Lam,B., Southwick,A., Miramanda,M., Brooks,S.,
BUEHLER,E., Chou,Q., Chin,C., Chlou,J., Choi,E., Dunn,P., Gonzalez,A.,
Howng,B., Kim,C., KOO,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
MUKHARSKY,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
THAVERI,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A. and Davis,R.W.
Direct Submission
Submitted (28-MAR-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
3 (bases 1 to 102183)
FEDERSPIEL,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
ALTAFI,H., Nguyen,M., Lam,B., Southwick,A., Miramanda,M., Brooks,S.,
BUEHLER,E., Chou,Q., Chin,C., Chlou,J., Choi,E., Dunn,P., Gonzalez,A.,
Howng,B., Kim,C., KOO,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
MUKHARSKY,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
THAVERI,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A. and Davis,R.W.
Direct Submission
Submitted (21-JUN-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
4 (bases 1 to 102183)
FEDERSPIEL,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
ALTAFI,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.
and Davis,R.W.
Direct Submission
Submitted (15-AUG-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Jun 21, 2000 this sequence version replaced gi:8134868.
Bases 1-55,850 of F13B4 overlap with bases 47,308-103,157 of BAC
Clone T6J4, gb|AC011810 and bases 76,686-102,183 of F13B4 overlap
with bases 1-25,498 of BAC clone F21F23, gb|AC027656.
e-mail for correspondence: arabsequence.stanford.edu
Genes with
similarity to proteins in the databases are named 'putative',
'-like' or 'similar to'. Genes that have EST similarity but no
significant protein similarity are described as 'unknown proteins'.
Genes that are annotated based only on gene prediction software
are described as 'hypothetical proteins'. The gene prediction
programs used to predict genes include: Grail (Informatics Group,
Oak Ridge National Laboratory,
http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
http://genes.mit.edu/GENSCAN.), FEXA (Victor Solovyev,
http://genomic.sanger.ac.uk/gf/gf.shtml), and NetPlantGene (S.M.
Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).
FEATURES
source
1..102183
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="I"
/clones="F13B4"
1..55850
/sequence="overlap with bases 47,308-103,157 of BAC clone
T6J4, gb|AC011810. See GenBank record for clone T6J4 for
annotation in this region."
misc_feature
1..55850
/sequence="overlap with bases 47,308-103,157 of BAC clone
T6J4, gb|AC011810. See GenBank record for clone T6J4 for
annotation in this region."
gene
/sequence="F13B4.1"
/sequence="F13B4.1"
/sequence="Hypothetical protein"
/codon_start=1
/protein_id="AAF99819.1"
/db_xref="GI:9802750"

```

```

Howng,B., Johnson-Hopson,C., Khan,S., Kim,C., Koo,T., Lee,J.M.,
Lenz,C., Liu,A., Liu,S., Mukharasy,N., Pham,P., Sakano,H.,
Shinn,P., Toriumi,M., Vaysberg,M., Yu,G., Ecker,J., Theologis,A.
and Davis,R.W.
Unpublished
2 (bases 1 to 102183)
FEDERSPIEL,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
ALTAFI,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,
Chin,C., Chlou,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C.,
KOO,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
MUKHARSKY,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
THAVERI,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A. and Davis,R.W.
Direct Submission
Submitted (28-MAR-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
3 (bases 1 to 102183)
FEDERSPIEL,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
ALTAFI,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,
Chin,C., Chlou,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C.,
KOO,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
MUKHARSKY,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
THAVERI,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A. and Davis,R.W.
Direct Submission
Submitted (21-JUN-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
4 (bases 1 to 102183)
FEDERSPIEL,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
ALTAFI,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.
and Davis,R.W.
Direct Submission
Submitted (15-AUG-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Jun 21, 2000 this sequence version replaced gi:8134868.
Bases 1-55,850 of F13B4 overlap with bases 47,308-103,157 of BAC
Clone T6J4, gb|AC011810 and bases 76,686-102,183 of F13B4 overlap
with bases 1-25,498 of BAC clone F21F23, gb|AC027656.
e-mail for correspondence: arabsequence.stanford.edu
Genes with
similarity to proteins in the databases are named 'putative',
'-like' or 'similar to'. Genes that have EST similarity but no
significant protein similarity are described as 'unknown proteins'.
Genes that are annotated based only on gene prediction software
are described as 'hypothetical proteins'. The gene prediction
programs used to predict genes include: Grail (Informatics Group,
Oak Ridge National Laboratory,
http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
http://genes.mit.edu/GENSCAN.), FEXA (Victor Solovyev,
http://genomic.sanger.ac.uk/gf/gf.shtml), and NetPlantGene (S.M.
Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).
FEATURES
source
1..102183
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="I"
/clones="F13B4"
1..55850
/sequence="overlap with bases 47,308-103,157 of BAC clone
T6J4, gb|AC011810. See GenBank record for clone T6J4 for
annotation in this region."
misc_feature
1..55850
/sequence="overlap with bases 47,308-103,157 of BAC clone
T6J4, gb|AC011810. See GenBank record for clone T6J4 for
annotation in this region."
gene
/sequence="F13B4.1"
/sequence="F13B4.1"
/sequence="Hypothetical protein"
/codon_start=1
/protein_id="AAF99819.1"
/db_xref="GI:9802750"

```

gene  
CDS  
/translation="MYVTKRLSEYQKNPSELTSLTEGPNSGVLVIQDEESRTPCCFG  
KCFDWMKGLPFPNTKTVRFQGVKRVYLDPAVFIPLVHLPLSLNLIYVIRQSGKQT  
GEACVATGEGTSCICFSPIDADRPDLDFDIYQQFPIHQSGSNLIYVIRQSGKQT  
EGIPQFLTRQWAAASLPFEYGLIDDAKGVVDKLRLEFPLDKSVVVGKWWYVFL  
FYKGEQDANDQNKMSYMTLQORPEEFVFCENTGNKHFFVVDVEVETVVKLEG  
DKIARETKGVNSDGVVVRSGTQKIGLGSVVLERMKWEERFSGSKDELRSIKRT  
EKFEQGGPHKSVRCVLTVELTKETDGLSLVTFEKKHVDKLSKWD"  
/gene="F13B4.2"  
complement (58592..61490)  
complement (join (58592..59438,59566..59665,59759..59905,  
60390..61084,61171..61490))  
/gene="F13B4.2"  
/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AAF99820.1"  
/db\_xref="GI:9802751"

gene  
CDS  
/translation="MYVTCLSEYQKNPSELTPTEGPNSGVLVIQDHSQTRPCCF  
GLCLVDSYLSGLPLPQNLKLAARFVAGDDATCDPVFIPVNLKPLSSNLIYVIRQ  
KXHGAPASGKEEDRVPCCFSSVYPEAKPQLDPYDVIYQFPMHLKXSSYVYAT  
SVAPDVSFWFLKKNWTEYSRFQDFELDKDAKGLNTELRTELGLGNSVVGKVV  
PFIVKRDVDOIKKSMYISMTLEQREWEVSIEKDKSRNTDVLVDVELDTEVVKV  
GGEISRGVANGFVAVGDKKIGLGSVVRMKWEERFGMTGKDNERSVMAKRL  
KSTDEGPNGLVYIQDEESRPTLKCFVNSFDWELNGLPQNTTVIRQIGTALYL  
DPAVFPVLHQPPSSNLYVIRGSKTGTEACVSAKEGDRASCFTYIPEAVRPL  
DPDINQOPEIHQGRSSTKKFPATSVASDGIPOPLTRKGSVIFSPAVEVDGLID  
AKGIVDAKLRLEPOLDNSVVGKVVYVFLVKEGDKADOMYKSMYNTLQOREEV  
FCFENVCKNRREVVDVEVEVHVKLEGERARETQGVNSDGVVWFNASKGKGLG  
SVLLEKMKWEERFGMTGKDNERSVMAKRL  
complement (62660..65932)  
/gene="F13B4.3"  
complement (join (62660..63503,63707..64040,64557..65380,  
65510..65621,65725..65932))  
/gene="F13B4.3"  
/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AAF99821.1"  
/db\_xref="GI:9802752"

gene  
CDS  
/translation="MYVTKRLSEYQKNPSELTPTEGPNSGVLVIQDEESQPTCCFG  
KCFGDLNGFPFKEEDRVPCCFSSVYPEAKPQLDPYDVIYQFPMHLKXSSYVYAT  
KXHGAPASGKEEDRVPCCFSSVYPEAKPQLDPYDVIYQFPMHLKXSSYVYAT  
SVASDGIPIRFLTRQWAAASLPFEYGLIDDAKGVVDKLRLEFPLDKSVVVGKWWYVFL  
VFLFYKGEQDANDQNKMSYMTLQORPEEFVFCENTGNKHFFVVDVEVETVVKLE  
EGENIGRETGVNSDGVVVRSGTQKIGLGSVVLERMKWEERFSGSKDELRSIKRT  
KXTEKEGKGVNSDGVVVRSGTQKIGLGSVVLERMKWEERFSGSKDELRSIKRT  
LSYQNSDLTSLPEGPNSGVLVIQDEESRPTLKCFVNSFDWELNGLPQNTTVIRQIGTALYL  
TYRTGDNRSRHDQVLTPIVLDQPIYNSRYVIRKQSGGASASAKEDRVPCC  
FCFSYVFPQADPDIYQFVHQKRPSSGYITTVAPNGIPIPEFLKRYTVE  
YNSQDFGLTDDAGINTLRSKLPNDMNTSVVVGKVVYVFIPIVFERDAKQIKSPY  
YSMTLQOREEVYSCVNVNKGGEVVDVETQVVKLEGOETISRETSGGFVWF  
SVLGDQKQKIGLGSVVRMKWEERFGMLNNGERSIKRSERFEGGSHWSYRC  
LVLIQSFELKMDGSLVITFEFTHVDKLSKWD"  
complement (66277..69118)  
/gene="F13B4.4"  
complement (join (66277..67099,67189..67554,67689..67786,  
67968..68724,68802..69118))  
/gene="F13B4.4"  
/note="Unknown protein"  
/codon\_start=1  
/protein\_id="AAF99822.1"  
/db\_xref="GI:9802753"

gene  
CDS  
/translation="MYVTCLSEYQKNPSELTPTEGPNSGVLVIQDEESLTCCFG  
LCNEIFLRLGFLPFRQWAAASLPFEYGLIDDAKGVVDKLRLEFPLDKSVVVGKWWYVFL  
YKXGESSSTKEVERVTCFCIRFNHVMYKPPQDPYDVIYQFPMHLKXSSYVYAT  
FAKVPALLPFLMRQGTIENSTLQDFGLIDDAKGLNLELSKLPGLDISVV  
KXNVPELIFYKEDGIDQVLSMYNTWTLHQREVEVFFVENVNHEBCVEVDVDEL  
QVNIQGVKSITINDDAKGVVWFQVQDKKIGLRSVVRMKWEERFGMLNNGERSIKRSERFEGGSHWSYRC  
VYTKLRDFRFGGSHWSYRCVLSWSVEYFWMQSGVMIQORGEDFVYLITLDERI  
KFFSKGTRFSETMYTIRSEYRNPALTPPPEGPNSGVLVIQDHSQTRPCCF  
GSLVVDLSGLPLPQNLKLAARFVAGDDATCDPVFIPVNLKPLSSNLIYVIRQSGKQT  
KXHGAPASGKEEDRVPCCFSSVYPEAKPQLDPYDVIYQFPMHLKXSSYVYAT  
SVAPDVSFWFLKKNWTEYSRFQDFELDKDAKGLNTELRTELGLGNSVVGKVV  
PFIVKRDVDOIKKSMYISMTLEQREWEVSIEKDKSRNTDVLVDVELDTEVVKV  
GGEISRGVANGFVAVGDKKIGLGSVVRMKWEERFGMTGKDNERSVMAKRL  
KSTDEGPNGLVYIQDEESRPTLKCFVNSFDWELNGLPQNTTVIRQIGTALYL  
DPAVFPVLHQPPSSNLYVIRGSKTGTEACVSAKEGDRASCFTYIPEAVRPL  
DPDINQOPEIHQGRSSTKKFPATSVASDGIPOPLTRKGSVIFSPAVEVDGLID  
AKGIVDAKLRLEPOLDNSVVGKVVYVFLVKEGDKADOMYKSMYNTLQOREEV  
FCFENVCKNRREVVDVEVEVHVKLEGERARETQGVNSDGVVWFNASKGKGLG  
SVLLEKMKWEERFGMTGKDNERSVMAKRL  
complement (62660..65932)  
/gene="F13B4.3"  
complement (join (62660..63503,63707..64040,64557..65380,  
65510..65621,65725..65932))  
/gene="F13B4.3"  
/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AAF99821.1"  
/db\_xref="GI:9802752"

gene  
mRNA  
CDS  
/translation="MGYIGAGVAAALHRVYKSGVDHSYLAKVLOPFWTRFVVKVPLW  
WPNMILMGEMFLVTSLSILGVYSPQSDSPPRMVHFAHGLLLFLYCTFFDAVDSKQA  
RTNSSPGLBLFDHGCALACAFKAMAFSTAMCGROTFNFWLSALIPFYGATWEHY  
FNTLILVINGFTGLALIFVSHFTALVGAEMWAQLOSGIPFWSVFFVNEIQT  
RAVILYAFVAFVPTVAFNVTVNYSRNSGVLALAMLYFVVLGGVLIWYLS  
PINTIATYPLVLTGLAFGLVGRMILAHLCDEPKGLKTNMCSLALYLPALANAL  
TARLNAGVLPVLGLGICPTVSLYHFAVSITHEITELGICYCFRITRKEA"  
complement (73331..74733)  
/gene="F13B4.6"  
complement (join (73331..73621,73703..73861,73933..74733))  
/gene="F13B4.6"  
/note="Unknown protein"  
/codon\_start=1  
/protein\_id="AAF99824.1"  
/db\_xref="GI:9802755"

Query Match 47.1%; Score 223.4; DB 8; Length 102183;  
Best Local Similarity 69.2%; Pred. No. 1.1e-59;  
Matches 413; Conservative 0; Mismatches 61; Indels 123; Gaps 3;  
QY 1 ATGGCTCGTACGAGCAACTCCAGAAATCACCGAGGAAAGCTCCGAGGACTCTG 60  
Db 19325 ATGGCTCGTACGAGCAACTCCCGAGAAATCACCGAGGAAAGCTCCGAGGAG 19266  
QY 61 CTGCTTACCAAGCGCGGAGAAATCTGCCGACTACTTGGAGGAGTCAAGAAACCTCAC 120  
Db 19265 CTGCTTACCAAGCGCGGAGAAATCTGCCGACTACTTGGAGGAGTCAAGAAACCTCAC 19206  
QY 121 CGTTACCGTCCGGAACCGTCTCTTC-----148  
Db 19205 CGTTACCGTCCGGAACCGTCTCTTCGTTAACTTTCACGTTTCTTCTTTCGAATCA 19146  
QY 149 -----GTGAGATTCT 159  
Db 19145 CAATCTTCTCATCGTCTTCAATTTCTGATGACTATTGTTGATGATCATGATGATCTG 19086  
QY 160 AAATACCAAGAGCAAGAGTTGTTGATCCGTAACCTTCTTCTTCAAGCTCTTGTCTG 219  
Db 19085 AAATACCAAGAGCAAGAGTTGTTGATCCGTAACCTTCAAGCTCTTGTCTG 19026  
QY 220 GAAATCGCTCAAGATTACAGACGGA-----TCTGAGATTCAGAGCATCTC 266  
Db 19025 GAAATCGCTCAAGATTACAGAGTAAAGATTGTTGTTTCAATCATCGAAACCTTAGAT 18966  
QY 267 GGTGTTAGTCTTCAAGAGCTGCTGAA-----GCA 297  
Db 18965 ACTGATTAGTTATACCTACTTGTATGAGAAAGATTGATTAATAATTGGTGAANTTGA 18906  
QY 298 TATTTGACGATCTGAGATTCCAGAGCCATCGCGTGTAGTCTTCTTCAAGAGCTCGGAG 357



[illegible]



Query Match 47.1%; Score 223.4; DB 8; Length 103157;  
Best Local Similarity 69.2%; Pred. NO. 1.1e-59;  
Matches 413; Conservative 0; Mismatches 61; Indels 123; Gaps 3;

COMMENT

annotation of this entry and other sequences of chromosomes 3, 4  
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

**FEATURES**  
 source  
 1..96379  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /variety="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="5"  
 37579..40123  
 /gene="TSK6\_10"  
 complement(join(37579..37728,37825..37914,38003..38121,  
 38249..38348,38439..38535,38628..38691,38775..38907,  
 38990..39169,39247..39386,39988..40123))  
 /gene="TSK6\_10"  
 complement(join(37579..37728,37825..37914,38003..38121,  
 38249..38348,38439..38535,38628..38691,38775..38907,  
 38990..39169,39247..39386,39988..40123))  
 /gene="TSK6\_10"  
 note="similarity to protein kinase 1, Populus nigra,  
 EMBL:AB041503  
 Contains Prokaryotic membrane lipoprotein lipid attachment  
 site AA113-123;Protein kinases signatures and profile  
 AA116-138;Protein kinases signatures and profile  
 AA231-243"  
 /codon\_start=1  
 /product="ser/thr specific protein kinase-like protein"  
 /protein\_id="CAC03450.1"  
 /db\_xref="GI:9795154"  
 /db\_xref="GOA:Q9FY63"  
 /db\_xref="SPTREMBL:Q9FY63"  
 /translation="MKQIVITLVLLQAVYLHQSCTVMSLTTOESPQPSAFTPAL  
 PEGSTSAQVNSLSHSFSGFYGLNLSIHDFDKAEKVTNLSIWLKGVSLIDVNI  
 LEGTSFKFSNLLQGGGFCVYSATLENNISAAVKLDCANEDAAKEFKSEVELLSK  
 LQENITSLIGYSTNDFARFIVYELMPVNSLSHLSHSGQSALTPMRMKIALDVTR  
 GLEYLHPCHEPFIHRLDKSNLLDSNFNAKISDFGLAVVDGPKNKHLSGTGVY  
 APEYLLNQGTKEGVDYAFGVLLLELLGKKPVEKLAPGBCQSIITWMPYLTDTKL  
 PSVIDPALKDTMDLKLHYQVAVALCVQPEPSYRPLITDVLHSLIPLVPELGGTLE  
 TIKCASMDHC"  
 complement(37579..37728)  
 /gene="TSK6\_10"  
 /number=1  
 complement(37729..37824)  
 /gene="TSK6\_10"  
 /number=1  
 complement(37825..37914)  
 /gene="TSK6\_10"  
 /number=2  
 complement(37915..38002)  
 /gene="TSK6\_10"  
 /number=2  
 complement(38003..38121)  
 /gene="TSK6\_10"  
 /number=3  
 complement(38122..38248)  
 /gene="TSK6\_10"  
 /number=3  
 complement(38249..38348)  
 /gene="TSK6\_10"  
 /number=4  
 complement(38349..38438)  
 /gene="TSK6\_10"  
 /number=4  
 complement(38439..38535)  
 /gene="TSK6\_10"  
 /number=5  
 complement(38536..38627)  
 /gene="TSK6\_10"  
 /number=5  
 complement(38628..38691)  
 /gene="TSK6\_10"  
 /number=6  
 complement(38692..38774)

/gene="TSK6\_10"  
 /number=6  
 complement(38775..38907)  
 /gene="TSK6\_10"  
 /number=7  
 complement(38908..38989)  
 /gene="TSK6\_10"  
 /number=7  
 complement(38990..39169)  
 /gene="TSK6\_10"  
 /number=8  
 complement(39170..39246)  
 /gene="TSK6\_10"  
 /number=8  
 complement(39247..39386)  
 /gene="TSK6\_10"  
 /number=9  
 complement(39387..39987)  
 /gene="TSK6\_10"  
 /number=9  
 complement(39988..40123)  
 /gene="TSK6\_10"  
 /number=10  
 42133..46290  
 /gene="TSK6\_20"  
 complement(join(42133..42389,42567..42618,43012..43170,  
 43361..43438,43873..43944,44028..44356,44450..44600,  
 44714..44773,44889..45055,45380..45452,45542..45691,  
 45790..46009,46091..46290))  
 /gene="TSK6\_20"  
 complement(join(42133..42389,42567..42618,43012..43170,  
 43361..43438,43873..43944,44028..44356,44450..44600,  
 44714..44773,44889..45055,45380..45452,45542..45691,  
 45790..46009,46091..46290))  
 /gene="TSK6\_20"  
 /codon\_start=1  
 /product="putative protein"  
 /protein\_id="CAC03451.1"  
 /db\_xref="GI:9795155"  
 /db\_xref="SPTREMBL:Q9FY62"  
 /translation="MKPFAINTLVNETHAKLSFGVMESSEIEGSESTVITSPSR  
 RVRELLALCFVSFLSSSSPRFGSNEKFAANFVIFQSVEEAGGFQDFESFVTELVSCI  
 DSYLVNVALDANNELEVEIEVDEILKVLSSPQMDQVDVIDALSFHLKPKVTSKFADL  
 SSRQLQVEIVDFVEACNPRDMLSLCEALDAARCYHSASTSTPLLHGLSKVFIL  
 IQRRHYQLKVAVPIVLNLVKDISLETDVQVEDLFDKALGTASSIRDVSSKINNEEA  
 KVRCLCLVYIQTALISVSRDKAASCIPLVQLBPLTSCGLTHLGLITGNDTEKL  
 MSTVAGDDDEFITSPFDISLGASLFTCAKISHEVAEANAVALGSSVVDLQNPVKFW  
 QATGLKIILSSGDLMEFKHAEFLDITKGTSSQCNDQIDCSDYTFGIYATILQ  
 AVTLLINYPADRLKKTFFALKRVLSIDIPAPHRFDVIRALVITNSRSPSMALLGLV  
 KDSMSKSLQDTPCAAVDTHVILVELVLRPPQGPPLPPQSDAILAALNLYRFALL  
 FESRDKAICEAGKERSKVGSDILSKKNLEKAYKEWLLPLRLTLVSCSIAENLKEDHQ  
 ESSLDVGLLNPIELVYRGIELVEEKLKSH"  
 complement(42133..42389)  
 /gene="TSK6\_20"  
 /number=1  
 complement(42390..42566)  
 /gene="TSK6\_20"  
 /number=1  
 complement(42567..42618)  
 /gene="TSK6\_20"  
 /number=2  
 complement(42619..43011)  
 /gene="TSK6\_20"  
 /number=2  
 complement(43012..43170)  
 /gene="TSK6\_20"  
 /number=3  
 complement(43171..43360)  
 /gene="TSK6\_20"  
 /number=3  
 complement(43361..43438)  
 /gene="TSK6\_20"  
 /number=4

```

intron      complement(43439..43872)
            /gene="T5K6_20"
            /number=4
exon        complement(43873..43944)
            /gene="T5K6_20"
            /number=5
intron      complement(43945..44027)
            /gene="T5K6_20"
            /number=5
exon        complement(44028..44356)
            /gene="T5K6_20"
            /number=6
intron      complement(44357..44449)
            /gene="T5K6_20"
            /number=6
exon        complement(44450..44600)
            /gene="T5K6_20"
            /number=7
intron      complement(44601..44713)
            /gene="T5K6_20"
            /number=7
exon        complement(44714..44773)
            /gene="T5K6_20"
            /number=8

Query Match      46.5%; Score 220.4; DB 8; Length 96379;
Best Local Similarity 72.6%; Pred. No. 9.8e-59;
Matches 312; Conservative 0; Mismatches 91; Indels 27; Gaps 1;

QY 71 AGCGCGGAGGAAATCTGCGCGGACTACTGGAGGAGTCAAGAACTCAACCGTTACCGTTC 130
DB 24161 AGCGCGGAGGAAATCAGCACCAACCAACCGGAGGAGTCAAGAACTCAACCGTTACCGTTC 24102

QY 131 CGGAGACCTCGCTCTCTGAGATTCGTAATACCAAGAGGACAGAGTGTGATCC 190
DB 24101 CAGGAGACCTCGCTCTCTGAGATTCGTAATACCAAGAGGAGTCAAGATTTGATCC 24042

QY 191 GTAAATCTCTTTCAACGCTTGTTCGTGAATCGCTCAAGATTCAGAGCGGATCTCA 250
DB 24041 GCAAGCTTCTCTTCCAGCGTCTGTTTCGGGAATAGCTCAGGATTCAGGTAATTTA 23982

QY 251 GATTCAGAGCCATCGCGGTGTAGCTCTTCAAGAAGCT----- 288
DB 23981 TACTCATTCAGCGCGGATTAATCAATCTATAGGTTATGTTATGGAATTAGGTATACT 23922

QY 289 -----GCTGAAGCATATTTCAGGATCTGAGATTCAGAGCCATGCGGTGTTAGCTCTTC 343
DB 23921 AATTAGTTAATTTTCAGAGCGGATCTGAGTTTCAGAGTCACGCGGTGTTAGCTCTTC 23862

QY 344 AAGAAGCTCGGAGGATATTGTTGGTGGTTGTTTGAAGACACAAATCTGTGTGCCATTC 403
DB 23861 AAGAAGCGGCTGAGGCATACCTAGTTGGACTTTTCAGGAGTACTAATCTCTGCGCTATTC 23802

QY 404 ATGCAAGAGGTTAGATTAATGCTTAAGATCTTCAATTTGCAAGAGATTTGTTGGAG 463
DB 23801 ACCTCAAGAGGTTAGATCATGCTTCAAGACATTCATTTGGCTAGAGAAATTCGTGGAG 23742

QY 464 AGCGTGCTTA 473
DB 23741 AACGTGCTTA 23732

RESULT 11
ATT30N20/c
LOCUS      Arabidopsis thaliana DNA chromosome 5, BAC clone T30N20 (BSSA
DEFINITION project).
ACCESSION AL365234
VERSION   AL365234.1 GI:8979707
KEYWORDS  Arabidopsis thaliana (thale cress)
SOURCE    Arabidopsis thaliana
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 98473)
Bevan,M., Peters,S.A., van Staveren,M., Dirkse,W., Stiekema,W.,
Bancroft,I., Mewes,H.W., Rudd,S., Lemcke,K. and Mayer,K.F.X.
Unpublished
2 (bases 1 to 98473)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (06-JUN-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES
Location/Qualifiers
1..98473
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
1424..3476
/gene="T30N20_10"
complement(join(1424..1768,1854..1928,2027..2100,
2220..2306,2463..2621,2720..2818,3035..3162,3397..3476))
/gene="T30N20_10"
complement(join(1424..1768,1854..1928,2027..2100,
2220..2306,2463..2621,2720..2818,3035..3162,3397..3476))
/gene="T30N20_10"
/note="similarity to protein phosphatase 2C, alfalfa,
PIR:T09640
Contains Mitochondrial energy transfer proteins signature
AA246-255; Protein phosphatase 2C signature AA64-72"
/codon_start=1
/product="protein phosphatase 2C-like protein"
/protein_id="CAB96829.1"
/db_xref="GI:8979708"
/db_xref="GOA:Q9LEW5"
/db_xref="SPTREMBL:Q9LEW5"
/translation="MGYLDLALSYNSNPQTVEAPASGGGLSQNGKFSYCYASSAGKRS
SMEDFFETRIDINGEIVGLFGVFGHGCAAAAYVKRHLFSLNLTHPKFSIDTKSAI
TDVNHPTSEILLKSNENRDAGSTASTAILVGDRLVAVNGDSPAVISRGSKATVS
RDHKPDQSDERERTENAGCFVWAGVLAVSARGRDLLKQYVADPEIQEIKDITLE
FLILASGLWDVFSNEAAVAMVKEVEDFDSAKLVGEAKIKRGSADNLTCVVVFLEK
KSASSSHISSSSKEAKEMPPPLGLIALISSNEAKQVOIGSGNKPENVTNRKPDTSRST
DTLTLENSVTDKV"
complement(1424..1768)
/gene="T30N20_10"
/number=1
complement(1769..1853)
/gene="T30N20_10"
/number=1
complement(1854..1928)
/gene="T30N20_10"
/number=2
complement(1929..2026)
/gene="T30N20_10"
/number=2
complement(2027..2100)
/gene="T30N20_10"
/number=3
complement(2101..2219)
/gene="T30N20_10"
/number=3
complement(2220..2306)
/gene="T30N20_10"
/number=4
complement(2307..2462)
/gene="T30N20_10"
exon
intron
exon
intron
exon
intron
exon
intron
exon
intron

```



```

QY      404 ATGCAAGAGGGTTACGATATAGCTCAATGAGATGTTCAATTGCGACAGGATTCTGTGGAG 463
Db      80289 ACGCAAGAGGGTTACGATATAGCTCAATGAGATGTTCAATTGCGACAGGATTCTGTGGAG 80230

QY      464 ACGCGTGCCTTA 473
Db      80229 AACGTGCTTA 80220

RESULT 12
LOCUS   CEY49E10/c 125590 bp DNA linear INV 10-DEC-2003
DEFINITION Caenorhabditis elegans YAC Y49E10, complete sequence.
ACCESSION Z98866
VERSION   Z98866.1 GI:3217981
KEYWORDS HTG.
SOURCE   Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
          Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
          Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
AUTHORS none.
TITLE   Genome sequence of the nematode C. elegans: a platform for
        investigating biology. The C. elegans Sequencing Consortium
        Science 282 (5396), 2012-2018 (1998)
JOURNAL 99059613
MEDLINE 99059613
REMARK   The C.elegans Sequencing Consortium.
REFERENCE
AUTHORS Barlow K.
TITLE   Direct Submission
JOURNAL Submitted (28-AUG-1997) Nematode Sequencing Project, Sanger
        Institute, Hinxton, Cambridge CB10 1SA, England and Department of
        Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
        jesus@sanger.ac.uk or rw@nematoide.wustl.edu
COMMENT  On Jun 13, 1998 this sequence version replaced gi:2528901.
        Coding sequences below are predicted from computer analysis, using
        predictions from Genefinder (P. Green, U. Washington), and other
        available information.
        Current sequence finishing criteria for the C. elegans genome
        sequencing consortium are that all bases are either sequenced
        unambiguously on both strands, or on a single strand with both a
        dye primer and dye terminator reaction, from distinct subclones.
        Exceptions are indicated by an explicit note.
        For a graphical representation of this sequence and its analysis
        see: http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
        name=Y49E10
        IMPORTANT: This sequence is NOT necessarily the entire insert of
        the specified clone. It may be shorter because we only sequence
        overlapping sections once, or longer because we arrange for a small
        overlap between neighbouring submissions.
        This sequence is the entire insert of clone Y49E10. The true left
        end of clone Y11182 is at 36231 in this sequence. The true right
        end of clone Y75B8 is at 34585 in this sequence. The start of this
        sequence (1..107) overlaps with the end of sequence AL033514.
        The end of this sequence (125486..125590) overlaps with the start
        of sequence AL132904.

FEATURES
         Location/Qualifiers
            1..125590
               /organism="Caenorhabditis elegans"
               /mol_type="genomic DNA"
               /strain="Bristol N2"
               /db_xref="taxon:6239"
               /chromosome="III"
               /clone="Y49E10"
               complement(join(284..412,472..753))
               /gene="his-72"
               complement(join(284..412,472..753))
               /gene="his-72"
               /standard_name="Y49E10.6"
               /note="contains similarity to Pfam domain: PF00125 (Core
               histone H2A/H2B/H3/H4), Score=111.9, E-value=3.9e-30, N=1
               cDNA EST CEESA66F comes from this gene; cDNA EST CESAA19F
               comes from this gene; cDNA EST CEEST62F comes from this
               gene; cDNA EST CEESX85R comes from this gene; cDNA EST
               CEESG32RD comes from this gene; cDNA EST CESAA64F comes from
               this gene; cDNA EST CEESG32F comes from this gene; cDNA
               EST CEESQ74F comes from this gene; cDNA EST Yk83f1.5 comes
               from this gene; cDNA EST Yk61g10.5 comes from this gene;
               cDNA EST Yk108g7.3 comes from this gene; cDNA EST
               Yk80a4.3 comes from this gene; cDNA EST Yk63d8.3 comes
               from this gene; cDNA EST Yk376h11.3 comes from this
               gene; cDNA EST Yk376b11.3 comes from this gene; cDNA
               EST Yk402h8.3 comes from this gene; cDNA EST Yk458e6.3
               comes from this gene; cDNA EST Yk481h9.3 comes from this
               gene; cDNA EST Yk482e5.3 comes from this gene; cDNA
               EST Yk473h10.3 comes from this gene; cDNA EST Yk477a9.3
               comes from this gene; cDNA EST Yk477b9.3 comes from this
               gene; cDNA EST Yk477a9.3 comes from this gene; cDNA
               EST Yk477b9.3 comes from this gene; cDNA EST Yk477b11.3
               comes from this gene; cDNA EST Yk484d6.3 comes from this
               gene; cDNA EST Yk266a12.5 comes from this gene; cDNA
               EST Yk482e5.5 comes from this gene; cDNA EST Yk631a12.3
               comes from this gene; cDNA EST CEESG32R comes from this
               gene; cDNA EST Yk631a12.3 comes from this gene; cDNA
               EST Yk519d8.3 comes from this gene; cDNA EST Yk527g10.3
               comes from this gene; cDNA EST Yk527h4.3 comes from this
               gene; cDNA EST Yk586f3.3 comes from this gene; cDNA
               EST Yk592c11.3 comes from this gene; cDNA EST Yk606a8.3
               comes from this gene; cDNA EST Yk606b10.3 comes from
               this gene; cDNA EST Yk606g4.3 comes from this gene;
               cDNA EST Yk610g9.3 comes from this gene; cDNA EST
               Yk645f6.3 comes from this gene; cDNA EST Yk646a12.3
               comes from this gene; cDNA EST Yk652c12.3 comes from this
               gene; cDNA EST Yk652c7.3 comes from this gene; cDNA
               EST Yk658f11.3 comes from this gene; cDNA EST Yk666h11.3
               comes from this gene; cDNA EST Yk672a6.3 comes from this
               gene; cDNA EST Yk675e9.3 comes from this gene; cDNA
               EST Yk677e9.3 comes from this gene; cDNA EST Yk499b7.5
               comes from this gene; cDNA EST Yk519d8.5 comes from this
               gene; cDNA EST Yk527h4.5 comes from this gene; cDNA
               EST Yk527g10.5 comes from this gene; cDNA EST Yk586f3.5
               comes from this gene; cDNA EST Yk592c11.5 comes from
               this gene; cDNA EST Yk606a8.5 comes from this gene;
               cDNA EST Yk606b10.5 comes from this gene; cDNA EST
               Yk606g4.5 comes from this gene; cDNA EST Yk610g9.5
               comes from this gene; cDNA EST Yk631a12.5 comes from this
               gene; cDNA EST Yk637a12.5 comes from this gene; cDNA
               EST Yk645f6.5 comes from this gene; cDNA EST Yk646a12.5
               comes from this gene; cDNA EST Yk652c12.5 comes from
               this gene; cDNA EST Yk652c7.5 comes from this gene;
               cDNA EST Yk658f11.5 comes from this gene; cDNA EST
               Yk666h11.5 comes from this gene; cDNA EST Yk672a6.5
               comes from this gene; cDNA EST Yk675e9.5 comes from this
               gene; cDNA EST Yk677e9.5 comes from this gene; cDNA
               EST Yk510b11.5 comes from this gene; cDNA EST Yk201g11.3
               comes from this gene; cDNA EST Yk215g5.3 comes from
               this gene; cDNA EST Yk216b3.3 comes from this gene;

```

from this gene  
 cDNA EST yk228d6.3 comes from this gene; cDNA EST yk228h3.3 comes from this gene; cDNA EST yk246c5.3 comes from this gene  
 cDNA EST yk246f4.3 comes from this gene; cDNA EST yk266a12.3 comes from this gene; cDNA EST yk319d5.3 comes from this gene  
 cDNA EST yk330a10.3 comes from this gene; cDNA EST yk333g12.3 comes from this gene  
 /codon\_start=1  
 /product="C. elegans HIS-72 protein (corresponding sequence Y49E10.6)"  
 /protein\_id="CAB11546.1"  
 /db\_xref="GI:3979986"  
 /translation="MARTQTKARKSGKAPKQKATKAARKSAPTTGGVKPHRYRP  
 GTVALRIYRYOKSTELLRLKLPQRLVREIAQDFYDLRFQSAALGALQEAAYLV  
 GUFEDNLCAIAKRVITMPKQMLARRIRGERA"  
 join(3057..3182,3235..4008,4703..5053)  
 /genes="rpt-6"  
 join(3057..3182,3235..4008,4703..5053)  
 /genes="rpt-6"  
 /standard\_name="Y49E10.1"  
 /note="contains similarity to Pfam domain: PF00004 (ATPases associated with various cellular activities (AAA)), Score=327.8, E-value=4.1e-95, N=1  
 cDNA EST yk78f3.3 comes from this gene  
 cDNA EST yk91c7.3 comes from this gene  
 cDNA EST yk78f3.5 comes from this gene  
 cDNA EST yk91c7.5 comes from this gene  
 cDNA EST yk55a9.3 comes from this gene  
 cDNA EST yk56e6.3 comes from this gene  
 cDNA EST yk55a9.5 comes from this gene  
 cDNA EST yk56e6.5 comes from this gene  
 cDNA EST yk170d9.5 comes from this gene  
 cDNA EST yk205g7.5 comes from this gene  
 cDNA EST yk196b7.5 comes from this gene  
 cDNA EST yk231f1.5 comes from this gene  
 cDNA EST yk233f5.5 comes from this gene  
 cDNA EST yk319g11.5 comes from this gene  
 cDNA EST yk304d10.5 comes from this gene  
 cDNA EST yk361e3.5 comes from this gene  
 cDNA EST yk404d10.5 comes from this gene  
 cDNA EST yk453f1.5 comes from this gene  
 cDNA EST yk411g8.5 comes from this gene  
 cDNA EST yk440d10.5 comes from this gene  
 cDNA EST yk445c3.5 comes from this gene  
 cDNA EST yk512e1.5 comes from this gene  
 cDNA EST yk559a2.5 comes from this gene  
 cDNA EST yk567c5.5 comes from this gene  
 cDNA EST yk605h3.5 comes from this gene"  
 /codon\_start=1  
 /product="C. elegans RPT-6 protein (corresponding sequence Y49E10.1)"  
 /protein\_id="CAB11558.1"  
 /db\_xref="GI:3979998"  
 /translation="MAPPASTASSADPSKPTAQLTEESDEKTLRYFKYFSTKVDDAQQK  
 VADKSNVRRLQAPNELMTKRLKEELQQLHQEYSYGEVSKMDKKLVVKVHPK  
 KGVVDVDSIDINSLTARVALRADSLVHLKLNKVDPLVSLMWVLPDSTYEM  
 VGLDKQIKELVEPLVKVHPFLDGLTAQPKVLLPGPGCTGLLARAHAHTE  
 CTFIRVSGELVQKFGEGNRMVRELFVAREHAPSLIFWDEIDSLGSSRVGSSGGD  
 SEVQRTLELLNLDGFEATNKIKVIMATNRIDILDPALLRPGRIDRKKIEFFAPDEKA  
 RADILKTHKRNLMRGINNKIAEQIPGASGAEVKSCTEAGMFALRERRIHVQED  
 FENAVGVKMQDSEKKNISKKLWK"  
 complement(join(5233..5457,6234..6437))  
 /genes="Y49E10.2"  
 complement(join(5233..5457,6234..6437))

gene  
 CDS

Query Match 46.0%; Score 218.2; DB 3; Length 125590;  
 Best Local Similarity 67.9%; Pred. No. 5.2e-58;  
 Matches 321; Conservative 0; Mismatches 148; Indels 4; Gaps 1;

Db 753 ATGGCTCGTACCAAGCAAAACCGCTCGTAAATCCACGGAGGAAAGGCTCCAAAGAACGAG 694  
 QY 61 CTCGCTCAAGGCGCGAGGAAATCTGCGCGAGTACTGCGAGGAGTCAAGAAACCTCAC 120  
 Db 693 TTGGCCACCAAGGCGCGCAAAATCGGCTCCAAACACCGGAGGAGTCAAGAACCCACAT 634  
 QY 121 CGTTACCGTCCCGGAACCGTCTGCTGTTAGATTCGTTAAATACAGAAAGACACAGAG 180  
 Db 633 CGTTACCGTCCAGGAACGTGCTCTCGTGAGATTCGTTACAGAAAGTCGACTGAG 574  
 QY 181 TTGTTGATCCGPAATCTTCCTTTTCAACGTCTTGTGTTGAAATCGTCAAGATTACAAG 240  
 Db 573 CTCTCATCCGCAAGCTTCCTTTCCAAAGCTCTGTTCCGTGAGATTCGCCAGGACTTCAAG 514  
 QY 241 ACGGATCTGAGATTCAGAGCCATCGGCTGTTAGCTCTTCAAGAGCTCTCTAGACATAT 300  
 Db 513 ACTGATCTCCGCTTCAGTCGGCTGCCATCGGAGCTCTCC---AGGTTGTAGTTTGA 458  
 QY 301 TTGACGGATCTGAGATTCAGAGCCATCGGCTGTTAGCTCTTCAAGAAAGCTCGGAGGCA 360  
 Db 457 TTGAAATTTAAAAAATAATTAATTAATCTTTTTCAGGAGGCACTCGAAGCA 398  
 QY 361 TATTTGCTGGGTTGTTTGAAGACAAATCTGTGTGCCATTCATGCAAAAGAGGTTACG 420  
 Db 397 TACCTCGTCGAGCTCTTTCAGGAGCACCAACCTGTGCGCATCCACGCCAAGCGCTCAC 338  
 QY 421 ATAATGCCCTAAAGATCTTCAATTCGCAAGAGGATTCGTGGAGAGCGTCTTAA 473  
 Db 337 ATCATGCCAAAGACATGCAATTCGCCAGACCATCCGCGAGAGCGTCTTAA 285

RESULT 13  
 DHHS334  
 LOCUS D.hydei gene for histone H3.3B. 4534 bp DNA linear INV 30-AUG-1995  
 DEFINITION X81208  
 VERSION X81208.1 GI:963025  
 KEYWORDS histone H3.3.  
 SOURCE Drosophila hydei  
 ORGANISM Drosophila hydei  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1  
 AUTHORS Akhmanova, A.S., Bindels, P.C., Xu, J., Miedema, K., Kremer, H. and Hennig, W.  
 TITLE Structure and expression of histone H3.3 genes in Drosophila melanogaster and Drosophila hydei  
 JOURNAL Genome 38 (3), 586-600 (1995)  
 MEDLINE 96023949  
 PUBMED 7557364  
 REFERENCE 2 (bases 1 to 4534)  
 AUTHORS Akhmanova, A.S.  
 DIRECT SUBMISSION  
 TITLE Submitted (03-SEP-1994) A.S. Akhmanova, Catholic University of Nijmegen, Dept of Molecular & Developmental, Genetics, Toernooiveld, 6525 ED Nijmegen, NETHERLANDS  
 JOURNAL Location/Qualifiers  
 FEATURES  
 source  
 1..4534  
 /organism="Drosophila hydei"  
 /mol\_type="genomic DNA"  
 /strain="tubingen"  
 /db\_xref="taxon:7224"  
 /clone="Dhe32.1"  
 /clone\_lib="EMBL4"  
 295..3056  
 /gene="histone H3.3"  
 join(295..373,2568..2861,2928..3056)  
 /gene="histone H3.3"  
 295..373  
 /gene="histone H3.3"  
 /number=1  
 374..2567  
 intron

QY 1 ATGGCTCGTACGAGCAAACTCGAAGAAATCACCGAGGAAAGCTCCGAGGACTCTG 60

```

/gene="histone H3.3"
/number=1
2568..2861
/gene="histone H3.3"
/number=2
join(2580..2861,2928..3056)
/gene="histone H3.3"
/codon_start=1
/protein_id="CAA57081.1"
/db_xref="GI:963026"
/db_xref="FLYBASE:FBgn0014722"
/db_xref="GOA:P06351"
/db_xref="SWISS-PROT:P06351"
/translation="WARKQPKARKSTGKAPKOLATKARKAPSTGGVKKPHRYRP
GTVALREIRYQKTELLIRKLPFORLVREIAQPKDLRFQSAAGALQAPSEAYLV
GLFEDTLNCAIHKARVTIMPDKIQLARRIRGERA"
2862..2927
/gene="histone H3.3"
/number=2
2928..3056
/gene="histone H3.3"
/number=3

ORIGIN
Query Match 45.3%; Score 214.8; DB 3; Length 4534;
Best Local Similarity 68.5%; Pred. No. 3.8e-57;
Matches 326; Conservative 0; Mismatches 147; Indels 3; Gaps 2;

Qy 1 ATGGCTGTAGACAACTGCAAGAAATCACACGAGGAGAAAGTCCGAGACTCTG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2580 ATGGCAGTACCAAGCAGACGCCCGTAATCACTGGAGGCAAGGCTCTCGCAAGCA 2639
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 CTCGCTACCAAGCGGAGGAAATCTCGCGGACTACTGGAGGAGTCAAGAACCTCAC 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2640 CTGGCCACTAAGGCAGCCCGTAATCGGCCCGCTCCACAGGTGGTGTAAAGAGCCCCAT 2699
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 CGTTACCGTCCGGAACCGTCTCTCGTGAGATTCTGTAATACCAAGAGCACAGAG 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2700 CGTTATCGTCCCGTACTGTGGCTTGGTGAGATCCGTGCTACCAAGAAATCGACCGAG 2759
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 TTGTTGATCGGTAATCTCTTTTCAACAGTCTGTGTGTAATCGCTCAAGATTACAAG 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2760 TTGTCATCCGCAACTCGCTTTCACACGGTGTGGTGCGTGAATCGCAAGATTTCAG 2819
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 ACGGATCTGAGATCCAGAGCCATCGGTGTGTAGCTCTTCAAGAGCTGC--TGAAGCAT 298
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2820 ACCGATTTGGTTTCCATCGGCAGCTATCGGTGCTTGCAGTAAAGCAACAACT 2879
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
299 ATTTGACGGATCTGAGATTCAGAGCCATGCGGT-GTTAGCTCTTTCAGAGAGCTCGGAG 357
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2880 ATTTAACCCATTGGAATTTCAATAATACGAAATAACTAATGTTGCAGGAAGCCTCCGAG 2939
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
358 GCATTTTGGTGGTGTGTTTGAAGACACAATCTGTGTGCAATCTCAAGAGAGGTT 417
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2940 GCCTATTTGGTGGTGTATTTGAAGACACCAACTTTGTGCGCATTCACGCCAAGCGTGT 2999
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
418 ACGATAATGCCCTAAAGATGTTCAATTGCAAGAGGATTTCGTGGAGAGCGTGTCTTA 473
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3000 ACCATTATGCCAAGACATTCAGTTGGCGCTCGCATTCGGCGGAGGCTGTCTTA 3055

RESULT 14
AC130200
LOCUS AC130200 119578 bp DNA linear HTG 11-SEP-2003
DEFINITION Medicago truncatula clone mthl-64n13, WORKING DRAFT SEQUENCE.
ACCESSION AC130200
VERSION AC130200.28 GI:33147116
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

```

```

Medicago.
1 (bases 1 to 119578)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Medicago truncatula BAC Clone mthl-64n13
Unpublished
2 (bases 1 to 119578)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (08-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 119578)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (11-SEP-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Jul 23, 2003 this sequence version replaced gi:32567827.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 119578: contig of 119578 bp in length.
FEATURES
Location/Qualifiers
1..119578
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mthl-64n13"
/clone_lib="Medicago truncatula BAC library H1"
/note="This is one of two clones in the same well from
mthl-64n13"

ORIGIN
Query Match 44.3%; Score 210; DB 2; Length 119578;
Best Local Similarity 69.8%; Pred. No. 2.2e-55;
Matches 331; Conservative 0; Mismatches 80; Indels 63; Gaps 1;

Qy 1 ATGGCTGTAGAGCAAACTGCAAGAAATCACACGAGGAGAAAGCTCCGAGACTCTG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96853 ATGGCAGTACCAAAACAAACAGACACGCAAAATCCACCGCGGAAAGCCCAACAA 96912
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 CTCGCTACCAAGCGGAGGAAATCTCGCGGACTACTGGAGGAGTCAAGAAACCTCAC 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96913 CTCGCCACCAAGCGGAGGAAATCAGTCCGCGGAGGAGGAGTGAAGAGCCACAC 96972
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 CGTTACCGTCCGGAACCGTCTCTCGTGAGATTCTGTAATACCAAGAGAGCACAGAG 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96973 CGTTCCGTCGCGGACCGGTGGCGCTTCGTGAAATCCGGAATACCAAGAGACACTGAG 97032
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 TTGTTGATCCGTAACCTCTCTTTTCAACGCTCTCTGCTGAAATCGCTCAAGATTACAAG 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
97033 CTTCTGATTAAGAAATCTTCATTCCAACGCTCTGTTCTGAAATCGCTCAGATTTCAAA 97092
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 ACGGATCTGAGATTCCAGAGCCATCGCGTGTGTAGCTCTTCAAGAGAGTGTGTGAAGCATAT 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
97093 ACCGATCTCAGATTCCAGAGCAGTGTCTTCTCTCTCTTCAAGAGAGCGGTGAA----- 97146
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 TTGACCGATCTGAGATTCCAGAGCCATCGCGTGTGTAGCTCTTCAAGAGAGTTCGCGAGGCA 360
|||

```



Db 97147 -----GCT 97149

QY 361 TATTGGTGGTTGTTGAAGACACAAATCTGTGTCCTATCATGCAAGAGGGTTACG 420

Db 97150 TATCTTGTGTTGTTGTTGAAGATACAAATCTTTGCGCTATTCATGCTAAGAGAGTTACT 97209

QY 421 ATAATGCTAAAGATGTTCAATTGGCAGAAGGATTGCTGAGAGCGTCTTAG 474

Db 97210 ATTATGCTAAGATATCACTTGCAGAGAAATTCGTGTGAAGGGCTTAG 97263

RESULT 15

AY078027

LOCUS AY078027 411 bp mRNA linear PLN 18-FEB-2002

DEFINITION Arabidopsis thaliana AT4g40030/75J17\_200 mRNA, complete cds.

ACCESSION AY078027

VERSION AY078027.1 GI:18700235

KEYWORDS FTI CDNA.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 411)

AUTHORS Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.

TITLE Arabidopsis ORF clones

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 411)

AUTHORS Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.

TITLE Direct Submission

JOURNAL Submitted (04-FEB-2002) Salk Institute Genomic Analysis Laboratory (SIGnAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Shinn,P., Banh,J., Bowser,L., Chan,M.M., Chang,E., Dale,J.M., Deng,J.M., Goldsmith,A.D., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Davis,R.W., Theologis,A., and Ecker,J.R.

Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk) contributed equally to this work as PIs.

location/Qualifiers

1. .411

/organism="Arabidopsis thaliana"

/mol\_type="mRNA"

/db\_xref="taxon:3702"

/chromosome="4"

/clones="U12616"

FEATURES

source

1. .411

/ecotype="Columbia"

/notes="This clone is in pUNI 51"

1. .411

/notes="histone H3.3"

/codon\_start=1

/product="AT4g40030/75J17\_200"

/protein\_id="AAL77728.1" \_200"

/db\_xref="GI:18700235"

/translation="MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRRVP GTVALREIRKYQKSTELLIRKPPFQRLREIAODFKTLRFQSHVALQEAEEAIV GLFEDTNLCALHAKRVTIMEKDIQLARRINGERA"

ORIGIN

Query Match 44.0%; Score 208.6; DB 8; Length 411;

Best Local Similarity 79.4%; Pred. No. 2.6e-55;

Matches 247; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACAGGAGGAAAAGCTCCGAGGACTCTG 60

Db 1 ATGGCTCGTACGAAGCAAACTGCTGTAAGTCCACCGGAGGTAAGCTCCAGGAACAA 60

QY 61 CTGCTACCAAGGGGGAGGAAATCTGCCGCCGACTACTCGAGGAGTCAAGAACTCTAC 120

Db 61 CTTGCTACTAAGGCTGCTCGTAAATCTGCACCAACTACTCGTGGAGTCAAGAAACCA 120

QY 121 CGTTACCGTCCCGGAACCGTGGCTCTTCGTGAGATTCTGTAATACCAAGAGACACAG 180

Db 121 CGTTACCGTCTGGAACTGTTGCTCTCCGTGAATCCGTAAGTACCAAGAGTACTG 180

QY 181 TTGTTGATCGTAAACTTCCTTTTCAACGCTTTGTTGTTGTTGTTGTTGTTGTTGTT 240

Db 181 TTGCTTATCAGGAACCTGCCATTTTCAGAGGCTAGTCCGTGAGATTGCTCAAGATT 240

QY 241 ACGGATCTGAGATCCAGAGCCATGCGGTGTAGTCTTTCAGAAAGCTGCTGAAGATAT 300

Db 241 ACTGATTGCGTTTCCAGAGCCATGCTGTCTTAGTCTCTCCAGGAAGCTGCAGAGCAT 300

QY 301 TTGACGGGATCT 311

Db 301 CTTGTTGGTCT 311

Search completed: June 28, 2004, 16:23:04

Job time : 2085.37 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:24 ; Search time 1643.18 Seconds  
(without alignments)  
8614.203 Million cell updates/sec

Title: AAC50492

Perfect score: 474

Sequence: 1 ATGGCTGCTACGAGCAAAAC.....TTGCTGGAGCGTGCTTAG 474

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmv.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_man.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*  
29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	259.6	54.8	668	14	CD830068
2	256.4	54.1	572	14	CD820974
3	256.4	54.1	604	14	CD818139
4	256.4	54.1	622	14	CD834370

5	256.4	54.1	630	14	CD838405
6	256.4	54.1	635	14	CD842690
7	256.4	54.1	637	14	CD834027
8	256.4	54.1	639	14	CD838656
9	256.4	54.1	676	14	CD823357
10	255.4	53.9	527	9	AA650659
11	237.8	50.2	545	13	BQ791478
12	233	49.2	545	9	AV552968
13	233	49.2	603	14	CD830234
14	233	49.2	611	9	AI998977
15	233	49.2	643	14	CD822585
16	233	49.2	664	9	AV824078
17	232.6	49.1	414	13	BQ791662
18	231.4	48.8	593	14	CD829010
19	231.4	48.8	594	14	CD829250
20	231.4	48.8	599	14	CD830791
21	231.4	48.8	633	14	CD824194
22	231.4	48.8	634	14	CD819611
23	231.4	48.8	669	14	CD818523
24	230.2	48.6	606	9	AA713310
25	229.8	48.5	599	14	CD824297
26	229.8	48.5	605	14	CD819641
27	228.8	48.3	615	14	CD820205
28	228.8	48.3	689	14	CD839282
29	224.2	47.3	547	14	CD821103
30	223.6	47.2	453	9	AA713299
31	222.8	47.0	447	10	BE317214
32	222.8	47.0	501	10	BF643573
33	222.8	47.0	505	12	BI263311
34	222.8	47.0	571	10	AW574054
35	222.8	47.0	628	12	BG454253
36	222.8	47.0	630	12	BI263220
37	222.8	47.0	635	9	AJ496962
38	222.8	47.0	638	14	CA319769
39	222.8	47.0	656	12	BG453911
40	222.8	47.0	666	13	BQ165501
41	222.8	47.0	667	14	CF069311
42	222.8	47.0	670	12	BI271678
43	222.8	47.0	676	10	BE325005
44	222.8	47.0	678	12	BG455610
45	222.8	47.0	693	12	BI272111

#### ALIGNMENTS

RESULT 1  
CD830068

LOCUS

DEFINITION

BN40.044E03F011221 BN40 Brassica napus cDNA clone

sequence.

CD830068

CD830068.1 GI:32512008

EST.

Brassica napus (rape)

Brassica napus

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 668)

Genoplatte.

Genoplatte, a major partnership french program in plant genomics

Unpublished (2003)

Contact: Genoplatte

Genoplatte

93, rue Henri Rochefort 91025 EVRY CEDEX France

Tel: 33 1 69 47 54 00

Fax: 33 1 69 47 54 10

This sequence has been generated in the framework of the french

plant genomics programme 'Genoplatte' (<http://www.genoplatte.com>)

and <http://genoplatte-info.infobiogen.fr>.

Location/Qualifiers

source

1..668

CD830068 668 bp mRNA linear EST 10-JUL-2003  
BN40.044E03F011221 BN40 Brassica napus cDNA clone BN40044E03, mRNA

sequence.  
CD830068  
CD830068.1 GI:32512008  
EST.  
Brassica napus (rape)  
Brassica napus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 668)  
Genoplatte.  
Genoplatte, a major partnership french program in plant genomics  
Unpublished (2003)  
Contact: Genoplatte  
Genoplatte

93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme 'Genoplatte' (<http://www.genoplatte.com>)  
and <http://genoplatte-info.infobiogen.fr>.

Location/Qualifiers  
source  
1..668

```

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Jet neuf"
/db_xref="taxon:3708"
/clone="BN40044E03"
/tissue_type="seed"
/clone_lib="BN40"

```

## ORIGIN

```

Query Match      54.8%; Score 259.6; DB 14; Length 669;
Best Local Similarity 76.4%; Pred. No. 4.5e-66;
Matches 362; Conservative 0; Mismatches 49; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACCAAGCAAACTCGAGAAATCACAGAGGAGAAAGCTCCGAGGACTCTG 60
DB 122 ATGGCTCGTACCAAGCAGACTCGAGGAAATCACCGGAGGAGAAAGCTCCGAGGAGCAA 181
QY 61 CTCGCTTACCAAGCGCGGAGGAAATCTGCGCGAGTCTACTGAGGAGTCAAGAAACCTCAC 120
DB 182 CTCGCCACCAAGCTCGAGGAAATCACAGCCGACCCGAGGAGTCAAGAGGCTCAC 241
QY 121 CGTTACCGTCCCGGAAACCGTCTGCTGAGATTCGTAATATACAGAGGAGCAGAG 180
DB 242 CGTTACCGTCCCGGAAACCGTCTGCTGAGATTCGTAATATACAGAGGAGCAGAG 301
QY 181 TTGTTGATCGGTAAACCTTCTTTCAACGCTTGTGCTGAAATCGCTCAAGATTACAG 240
DB 302 TTACTGATCCGTAACCTTCTTTCCAGCGTCTGTGCTGAAATCGCCAGGATTTCAA- 360
QY 241 ACGGATCTGAGATTCAGAGCCATCGGCTGTAGTCTTCAAGAGCTGCTGAAGCATAT 300
DB 361 ----- 360
QY 301 TTGACGATCTGAGATTCAGAGCCATCGGCTGTAGTCTTCAAGAGCTGCGGAGGCA 360
DB 361 --GACGGATCTGAGATTCAGAGCCATCGGCTGTAGTCTTCAAGAGCTGCGGAGGCG 418
QY 361 TATTGTTGGTGGTGTGTTTGAAGACACAAATCTGTGTCGCAATCATGCAAGAGGGTTAG 420
DB 419 TACTTGGTGGTGTGTTTGAAGACACAAATCTGTGTCGCAATCATGCAAGAGTCAAC 478
QY 421 ATAATGCCCTAAAGATTTCAATTTGCAAGAGATTCGTTGGAGAGCGTCTTAG 474
DB 479 ATCATGCCCTAAAGATTTCAATTTGCTAGCGCATTCGTTGGAGAGCGTCTTAG 532

```

## RESULT 2

```

CD820974
LOCUS      BN25.040G21F011026 BN25 Brassica napus cDNA clone BN25040G21, mRNA
DEFINITION sequence.
ACCESSION CD820974
VERSION    CD820974.1 GI:32502914
KEYWORDS   EST.
SOURCE     Brassica napus (rape)
ORGANISM   Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 572)
            Genoplatte, a major partnership french program in plant genomics
            Genoplatte, a major partnership french program in plant genomics
            Unpublished (2003)
            Contact: Genoplatte
            Genoplatte
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplatte' (http://www.genoplatte.com)
            and http://genoplatte-info.inbioogen.fr.
FEATURES   Location/Qualifiers
            source          1..572

```

```

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Jet neuf"
/db_xref="taxon:3708"
/clone="BN25040G21"
/tissue_type="seed"
/clone_lib="BN25"

```

## ORIGIN

```

Query Match      54.1%; Score 256.4; DB 14; Length 572;
Best Local Similarity 75.9%; Pred. No. 3.8e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACCAAGCAAACTCGAGAAATCACAGAGGAGAAAGCTCCGAGGACTCTG 60
DB 56 ATGGCTCGTACCAAGCAGACTCGTAGGAAATCCACCGGAGGAGAAAGCTCCGAGGAGCAA 115
QY 61 CTCGCTTACCAAGCGCGGAGGAAATCTGCGCGAGTCTACTGAGGAGTCAAGAAACCTCAC 120
DB 116 CTCGCCACCAAGCGCGGAGGAAATCACAGCCGACCCGAGGAGTCAAGAGGCTCAC 175
QY 121 CGTTACCGTCCCGGAAACCGTCTGCTGAGATTCGTAATATACAGAGGAGCAGAG 180
DB 176 CGCTACCGTCCCGGAAACCGTCTGCTGAGATTCGTAATATACAGAGGAGCAGAG 235
QY 181 TTGTTGATCGGTAAACCTTCTTTCAACGCTTGTGCTGAAATCGCTCAAGATTACAG 240
DB 236 TTGTTGATCGGTAAACCTTCTTTCAACGCTTGTGCTGAAATCGCTCAAGATTACAG 294
QY 241 ACGGATCTGAGATTCAGAGCCATCGGCTGTAGTCTTCAAGAGCTGCTGAAGCATAT 300
DB 295 ----- 294
QY 301 TTGACGATCTGAGATTCAGAGCCATCGGCTGTAGTCTTCAAGAGCTGCGGAGGCA 360
DB 295 --GACGGATCTGAGATTCAGAGCCATCGGCTGTAGTCTTCAAGAGCTGCGGAGGCG 352
QY 361 TATTGTTGGTGGTGTGTTTGAAGACACAAATCTGTGTCGCAATCATGCAAGAGGGTTAG 420
DB 353 TATCTGTTGGTGTGTTTGAAGACACAAATCTGTGTCGCAATCATGCAAGAGTCAAC 412
QY 421 ATAATGCCCTAAAGATTTCAATTTGCAAGAGATTCGTTGGAGAGCGTCTTAG 474
DB 413 ATCATGCCCTAAAGATTTCAATTTGCTAGCGCATTCGTTGGAGAGCGTCTTAG 466

```

## RESULT 3

```

CD818139
LOCUS      BN20.044G19F011127 BN20 Brassica napus cDNA clone BN20044G19, mRNA
DEFINITION sequence.
ACCESSION CD818139
VERSION    CD818139.1 GI:32500079
KEYWORDS   EST.
SOURCE     Brassica napus (rape)
ORGANISM   Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 604)
            Genoplatte, a major partnership french program in plant genomics
            Genoplatte, a major partnership french program in plant genomics
            Unpublished (2003)
            Contact: Genoplatte
            Genoplatte
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplatte' (http://www.genoplatte.com)
            and http://genoplatte-info.inbioogen.fr.
FEATURES   Location/Qualifiers
            source          1..604

```

```
/organism="Brassica napus"  
/mol_type="mRNA"  
/cultivar="Jet neuf"  
/db_xref="taxon:3708"  
/clone="BN20044G19"  
/tissue_type="seed"  
/clone_lib="BN20"
```

## ORIGIN

Query Match	54.1%;	Score 256.4;	DB 14;	Length 604;
Best Local Similarity	75.9%;	Pred. No. 3.8e-65;		
Matches 360;	Conservative	0;	Mismatches 51;	Indels 63;
				Caps 1;

Qy	1	ATGGCTCGTACGAAGCAAACTGCAAGAAATACACGAGAGAAAGCTCCGAGGACTCTG	60
Db	62	ATGGCTCGTACGAAGCAACTGCTAGGAAATCCACCGAGAGAAAGCTCCGAGGAACAA	121
Qy	61	CTGCTACCAAGGCGGAGAGAAATCTCGCGGACTACTGAGGAGAGTCAAGAAACCTCAC	120
Db	122	CTGCCACCAAGGCGGCGAGAAATCGGCGCGACCACGGAGAGAGTCAAGAGGCTCAC	181
Qy	121	CGTTACCGTCCCGAAACGGTCGCTCTTCGTGAGATTCTGTAATACCAAGAGCAACAG	180
Db	182	CGTTACCGTCCCGAAACGGTCGCTCTTCGTGAGATTCTGTAATACCAAGAGCACTGAG	241
Qy	181	TTGTTGATCCGTAAACTCCTTTTCAACGCTCTGTTCTGTAATCGCTCAAGATTACAG	240
Db	242	TTGTTGATCCGTAAAGTTCGGTTTCAGCGTCTCGTTCTGTAGATCCGCCAGGATTTCAA	300
Qy	241	ACGGATCTGAGATTCCAGAGCCATGCGGTGTAGCTCTTCAAGAGAGCTGCTGAAGCATAT	300
Db	301	-----	300
Qy	301	TTGACGATCTGAGATTCCAGAGCCATGCGGTGTAGCTCTTCAAGAGAGCTCGGAGGCA	360
Db	301	--GACAGACCTTGAGGTTCCAGAGCCACGACGTGTGGCGCTTCAGGAAAGCTCGGAGGCT	358
Qy	361	TATTTGGTGGGTTGTTTGAAGACACAATCTGTGTGCATTCATGCAAGAGGGGTAGC	420
Db	359	TATCTGGTGGGTTTGTTCGAGACACAGAAATCTCTGTGCCATTCACGCCAGAGAGTCAAC	418
Qy	421	ATAAGCTAAAGATGTTCAAATTCGCAAGAGGATTCGTGGAGAGCGTGCTTAG	474
Db	419	ATCATGCTAAAGATTAATCAATTCGCTAGGCGTATTCGTGAGAGCGTGCTTAG	472

RESULT 4	CD834370	LOCUS	622 bp	mRNA	linear	EST 10-JUL-2003
DEFINITION	BN45.041116F011207	BN45 <i>Brassica napus</i> cDNA clone BN45041116, mRNA sequence.				

```

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Jet neuf"
/db_xref="taxon:3708"
/clone="BN45041116"
/tissue_type="seed"
/clone Tib="BN45"

```

## ORIGIN

Query Match	54.1%;	Score 256.4;	DB 14;	Length 622;
Best Local Similarity	75.9%;	Pred. No. 3.9e-65;		
Matches 360; Conservative	0;	Mismatches 51;	Indels 63;	Gaps 1;

Qy	1	ATCGCTCGTACGAAGCAAACTGCMAGAAAATCACACGGAGAAAAGCTCCGAGGACTCTG	60
Db	63	ATCGCTCGTACGAAGCAAGACTGTGTAGAAAATCCACCGGAGAAAAGCTCCCGAGGAGCAA	122
Qy	61	CTCGCTACCAAGCGCGGAGGAAATCTGCGCGGACTACTGAGAGAGTCAAGAAACCTCAC	120
Db	123	CTGCCACCAAGCGCGGCGAGGAAATCGCGCGCGACCCGAGAGAGATCAAGAGGCTCAC	182
Qy	121	CGTTACCGTCCCGGAAACCGTGCCTCTTCGTGAGATTCGTAAATACCAAGAGCACAGAG	180
Db	183	CGTTACCGTCCCGGAAACCGTGCCTCTTCGTGAGATTCGTAAATACCAAGAGCACACTCGAG	242
Qy	181	TTGCTGATCCGTAAACCTCTTTTCAACGCTCTGTCGTGAAATCGCTCAAGATTTACAG	240
Db	243	TTGTTGATCCGTAAAGCTTCGTTTTACGGCTTCGTTTCGTGAGATTCGCCCAGGATTTCAA	301
Qy	241	ACGGATCTGAGATTTCCAGAGCCATCGCGGTGTAGTCTCTTCAAGAGCTGCTGAAGCATAT	300
Db	302	-----	301
Qy	301	TTGACGGATCTGAGATTTCCAGAGCCATGCGGTGTAGTCTCTTCAAGAGCTGCGGAGGCA	360
Db	302	--GACAGACCTGAGGTTTCCAGAGCCACGCACTGTTGCGCGTCTCAGGAAGCTGCGAGGCT	359
Qy	361	TATTTGTTGGTTTGTTTGGAAGACACAAATCTGTGTGCCATTCATGTCAAGAAAGGGGTTCAG	420
Db	360	TATCTGTTGGTTTGTTCGAAGACACGAATCTCTGTGCCATTCACGCCAGAGAGTCAAC	419
Qy	421	ATAATGCTAAAGATGTTCAATTGGCAAGAGAGATTCGTGAGAGCGTGCTTTAG	474
Db	420	ATCATGCTAAAGATATTAATTGGCTAGGCGTATTCGTGAGAGCGTGCTTTAG	473

RESULT 5	CD838405	LOCUS	CD838405	630 bp	mRNA	linear	EST 10-JUL-2003
DEFINITION	RF02.111B20F010523	RF02	<i>Brassica napus</i>	cdna	clone	RF02111B20	mRNA sequence.

```

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="samourai (restored line)"
/db_xref="taxon:3708"
/clone="RFO211B20"
/tissue_type="anthers"
/clone_lib="RFO2"

```

## ORIGIN

```

Query Match      54.1%; Score 256.4; DB 14; Length 630;
Best Local Similarity 75.9%; Pred. No. 3.9e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACGAGCAAACTGCAAGAAATCAGAGGAGGAAAGCTCCGAGGACTCTG 60
DB 63 ATGGCTCGTACGAGCAGACTGTAGAAATCACCAGGAGGAAAGCTCCAGGAGCAA 122
QY 61 CTCGCTACCAAGCGCGAGGAAATCTGCCGCACTACTGGAGGAGTCAAGAAACCTCAC 120
DB 123 CTCGCCACCAAGCGCGAGGAAATCGGCCCGACCAACGGAGAGTCAAGAGCTCAC 182
QY 121 CGTTACCGTCCCGGAAACCGTCGCTCTTCGAGGATTCGTAATACCAAGAGACACAG 180
DB 183 CGTTACCGTCCCGGAAACCGTCGCTCTTCGAGGATTCGTAATACCAAGAGACACAG 242
QY 181 TTGTTGATCGTAAACTCTCTTCAACGCTCTGTTGTAATCGTGAATCGTCAAGATTACA 240
DB 243 TTGTTGATCGTAAACTCTCTTCAACGCTCTGTTGTAATCGTGAATCGTCAAGATTACA - 301
QY 241 ACGGATCTGAGATTCCAGAGCCATGCGGTGTAGTCTTCAAGAGCTGCTGAAGCATAT 300
DB 302 ----- 301
QY 301 TTGACGATCTGAGATTCAGAGCCATGCGGTGTAGTCTTCAAGAGCTGCGAGGCA 360
DB 302 --GACAGACTGAGGTTCAGAGCCATGCGGTGTAGTCTTCAAGAGCTGCGAGGCT 359
QY 361 TATTGCTGGTGTGTTGTTGAAGACACAAATCTCTGTCGCAATTCATCAAGAGGTTAG 420
DB 360 TATCTGTTGGTGTGTTGTTGAAGACACAAATCTCTGTCGCAATTCATCAAGAGGTTAG 419
QY 421 ATATGCTTAAAGATTTCAATTTGCAAGAGGATTCGTTGAGAGCGTCTTAG 474
DB 420 ATCATGCTTAAAGATTTCAATTTGCTAGGCGTATTCTGTTGAGAGCGTCTTAG 473

```

## RESULT 6

```

CD842690
LOCUS      RFO2.128J10F010914 RFO2 Brassica napus cDNA clone RFO2128J10, mRNA
DEFINITION
sequence.
ACCESSION  CD842690
VERSION     CD842690.1 GI:32524630
KEYWORDS   EST.
SOURCE     Brassica napus (rape)
ORGANISM   Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 635)
            Genoplatte.
            Genoplatte, a major partnership french program in plant genomics
            Unpublished (2003)
            Contact: Genoplatte
            Genoplatte
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplatte' (http://www.genoplatte.com
            and http://genoplatte-info.inbioogen.fr).
            Location/Qualifiers
            source          1. .635
FEATURES
            source

```

```

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="samourai (restored line)"
/db_xref="taxon:3708"
/clone="RFO2128J10"
/tissue_type="anthers"
/clone_lib="RFO2"

```

## ORIGIN

```

Query Match      54.1%; Score 256.4; DB 14; Length 635;
Best Local Similarity 75.9%; Pred. No. 3.9e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACGAGCAAACTGCAAGAAATCAGAGGAGGAAAGCTCCGAGGACTCTG 60
DB 63 ATGGCTCGTACGAGCAGACTGTAGAAATCACCAGGAGGAAAGCTCCAGGAGCAA 122
QY 61 CTCGCTACCAAGCGCGAGGAAATCTGCCGCACTACTGGAGGAGTCAAGAAACCTCAC 120
DB 123 CTCGCCACCAAGCGCGAGGAAATCGGCCCGACCAACGGAGAGTCAAGAGCTCAC 182
QY 121 CGTTACCGTCCCGGAAACCGTCGCTCTTCGAGGATTCGTAATACCAAGAGACACAG 180
DB 183 CGTTACCGTCCCGGAAACCGTCGCTCTTCGAGGATTCGTAATACCAAGAGACACAG 242
QY 181 TTGTTGATCGTAAACTCTCTTCAACGCTCTGTTGTAATCGTGAATCGTCAAGATTACA 240
DB 243 TTGTTGATCGTAAACTCTCTTCAACGCTCTGTTGTAATCGTGAATCGTCAAGATTACA - 301
QY 241 ACGGATCTGAGATTCCAGAGCCATGCGGTGTAGTCTTCAAGAGCTGCTGAAGCATAT 300
DB 302 ----- 301
QY 301 TTGACGATCTGAGATTCAGAGCCATGCGGTGTAGTCTTCAAGAGCTGCGAGGCA 360
DB 302 --GACAGACTGAGGTTCAGAGCCATGCGGTGTAGTCTTCAAGAGCTGCGAGGCT 359
QY 361 TATTGCTGGTGTGTTGTTGAAGACACAAATCTCTGTCGCAATTCATCAAGAGGTTAG 420
DB 360 TATCTGTTGGTGTGTTGTTGAAGACACAAATCTCTGTCGCAATTCATCAAGAGGTTAG 419
QY 421 ATATGCTTAAAGATTTCAATTTGCAAGAGGATTCGTTGAGAGCGTCTTAG 474
DB 420 ATCATGCTTAAAGATTTCAATTTGCTAGGCGTATTCTGTTGAGAGCGTCTTAG 473

```

## RESULT 7

```

CD834027
LOCUS      BN45.040F17F011019 BN45 Brassica napus cDNA clone BN45040F17, mRNA
DEFINITION
sequence.
ACCESSION  CD834027
VERSION     CD834027.1 GI:32515967
KEYWORDS   EST.
SOURCE     Brassica napus (rape)
ORGANISM   Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 637)
            Genoplatte.
            Genoplatte, a major partnership french program in plant genomics
            Unpublished (2003)
            Contact: Genoplatte
            Genoplatte
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplatte' (http://www.genoplatte.com
            and http://genoplatte-info.inbioogen.fr).
            Location/Qualifiers
            source          1. .637
FEATURES
            source

```

```

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Jet neuf"
/db_xref="taxon:3708"
/clone="BN45040R17"
/tissue_type="seed"
/clone_lib="BN45"

```

## ORIGIN

```

Query Match      54.1%; Score 256.4; DB 14; Length 637;
Best Local Similarity 75.9%; Pred. No. 3.9e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACACGGAGAAAGCTCCGAGGACTCTG 60
Db 63 ATGGCTCGTACGAAGCAAGACTGCTAGGAAATCCACCGGAGGAAAGCTCCGAGGACAA 122
QY 61 CTGCTACCAAGCGGCGGAGAAATCTGCGCGACTACTGAGGAGTCAAGAACTCTCAC 120
Db 123 CTGCTACCAAGCGGCGGAGAAATCTGCGCGACTACTGAGGAGTCAAGAACTCTCAC 182
QY 121 CGTTACCGTCCCGAAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGACACAGAG 180
Db 183 CGTTACCGTCCCGAAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGACACAGAG 242
QY 181 TTGTTGATCCGTAACTTCCTTTTCAACGTCCTTTCGTGAAATCGCTCAAGATTACAAG 240
Db 243 TTGTTGATCCGTAACTTCCTTTTCAACGTCCTTTCGTGAAATCGCTCAAGATTACAAG 301
QY 241 ACGGATCTGAGATTCAGAGCCATGCGGTGTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
Db 302 ----- 301
QY 301 TTGACGATCTGAGATTCAGAGCCATGCGGTGTAGCTCTTCAAGAGCTGCGAGGCA 360
Db 302 --GACAGACCTGAGTTCAGAGCCACGACGTGTTGGCGTTCAGGAGCTGCGAGGCT 359
QY 361 TATTGTTGGTGTGTTTGAAGACACAAATCTGTGTCCTATTCATGCAAGAGGTTACG 420
Db 360 TATTGTTGGTGTGTTTGAAGACACAAATCTGTGTCCTATTCATGCAAGAGGTTACG 419
QY 421 ATAATGCTTAAAGATGTTCAATTGGCAAGAGGATTCGTGAGAGCGTGTCTTAG 474
Db 420 ATCATGCTTAAAGATGTTCAATTGGCTAGGCTATTCTGTTGAGAGCGTGTCTTAG 473

```

```

RESULT 8
CD838656
LOCUS
DEFINITION
CD838656 639 bp mRNA linear EST 10-JUL-2003
sequence.
ACCESSION
CD838656.1 GI:32520596
VERSION
EST.
KEYWORDS
SOURCE
Brassica napus (rape)
ORGANISM
Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
1 (bases 1 to 639)
AUTHORS
Genoplante, a major partnership french program in plant genomics
TITLE
Unpublished (2003)
JOURNAL
COMMENT
Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.inbio.gen.fr).
LOCATION/Qualifiers
1. .639
FEATURES
source

```

```

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="samourai (restored line)"
/db_xref="taxon:3708"
/clone="RFO2112B20"
/tissue_type="anthers"
/clone_lib="RFO2"

```

## ORIGIN

```

Query Match      54.1%; Score 256.4; DB 14; Length 639;
Best Local Similarity 75.9%; Pred. No. 3.9e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACACGGAGAAAGCTCCGAGGACTCTG 60
Db 62 ATGGCTCGTACGAAGCAGACTGCTAGGAAATCCACCGGAGGAAAGCTCCGAGGACAA 121
QY 61 CTGCTACCAAGCGGCGGAGAAATCTGCGCGACTACTGAGGAGTCAAGAACTCTCAC 120
Db 122 CTGCTACCAAGCGGCGGAGAAATCTGCGCGACTACTGAGGAGTCAAGAACTCTCAC 181
QY 121 CGTTACCGTCCCGAAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGACACAGAG 180
Db 182 CGTTACCGTCCCGAAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGACACAGAG 241
QY 181 TTGTTGATCCGTAACTTCCTTTTCAACGTCCTTTCGTGAAATCGCTCAAGATTACAAG 240
Db 242 TTGTTGATCCGTAACTTCCTTTTCAACGTCCTTTCGTGAGATTCGCGGAGGATTTCAA 300
QY 241 ACGGATCTGAGATTCAGAGCCATGCGGTGTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
Db 301 ----- 300
QY 301 TTGACGATCTGAGATTCAGAGCCATGCGGTGTAGCTCTTCAAGAGCTGCGAGGCA 360
Db 301 --GACAGACCTGAGTTCAGAGCCACGACGTGTTGGCGTTCAGGAGCTGCGAGGCT 358
QY 361 TATTGTTGGTGTGTTTGAAGACACAAATCTGTGTCCTATTCATGCAAGAGGTTACG 420
Db 359 TATTGTTGGTGTGTTTGAAGACACAAATCTGTGTCCTATTCATGCAAGAGGTTACG 418
QY 421 ATAATGCTTAAAGATGTTCAATTGGCAAGAGGATTCGTGAGAGCGTGTCTTAG 474
Db 419 ATCATGCTTAAAGATGTTCAATTGGCTAGGCTATTCTGTTGAGAGCGTGTCTTAG 472

```

```

RESULT 9
CD823357
LOCUS
DEFINITION
CD823357 676 bp mRNA linear EST 10-JUL-2003
sequence.
ACCESSION
CD823357.1 GI:32505297
VERSION
EST.
KEYWORDS
SOURCE
Brassica napus (rape)
ORGANISM
Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
1 (bases 1 to 676)
AUTHORS
Genoplante, a major partnership french program in plant genomics
TITLE
Unpublished (2003)
JOURNAL
COMMENT
Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.inbio.gen.fr).
LOCATION/Qualifiers
1. .676
FEATURES
source

```

/organism="Brassica napus"  
/mol\_type="mRNA"  
/cultivar="Jet neuf"  
/db\_xref="taxon:3708"  
/clone="BN25048106"  
/tissue\_type="seed"  
/clone\_lib="BN25"

## ORIGIN

Query Match 54.1%; Score 256.4; DB 14; Length 676;  
Best Local Similarity 75.9%; Pred. No. 4e-65;  
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;  
QY 1 ATGGCTCGTACGAGCAAACTGCAGAAAATCACGGAGGAAAAGCTCCGAGGACTCTG 60  
DB 63 ATGGCTCGTACGAGCAGACTGCTAGAAAATCACCGGAGGAAAAGCTCCGAGGAGCAA 122  
QY 61 CTCGCTTACCAAGCGCGGAGGAAATCTGCGCGACTACTGGAGGAGTCAAGAAACCTCAC 120  
DB 123 CTCGCCACCAAGCGCGGAGGAAATCTGCGCGCCGACACCGGAGGAGTCAAGAGCTCAC 182  
QY 121 CATTACCGTCCCGGAACCGTCTCTTCGTGAGATTCTGTAATACCAAGAGACACAGAG 180  
DB 183 CGTTACCGTCCCGGAACCGTCTCTTCGTGAGATTCTGTAATACCAAGAGACACAGAG 242  
QY 181 TTGTTGATCGTAACTCTCTTTCAACGCTTGTTCGTGAATCGCTCAAGATTACAG 240  
DB 243 TTGTTGATCGTAACTCTCTTTCAACGCTTGTTCGTGAATCGCTCAAGATTACAG 301  
QY 241 ACGGATCTGAGATTCCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTGCTGAACATAT 300  
DB 302 ----- 301  
QY 301 TTGACGAGTCTGAGATTCCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTGCGAGGCA 360  
DB 302 --GACAGACTGAGTTCCAGAGCCAGCAGTCTTCGCGCTTCAAGAGCTGCGAGGCT 359  
QY 361 TATTGCTGGGTTGTTGAGGACACAACTCTGTGCGCAATTCATCAAGAGGGTTACG 420  
DB 360 TATCTGGTGGGTTGTTGAGGACACAACTCTGTGCGCAATTCATCAAGAGGGTTACG 419  
QY 421 ATAAATGCTAAAGATGTTCAATTGGCAAGAGGATTCTGTGAGAGCGTGTCTAG 474  
DB 420 ATCATGCTAAAGATGTTCAATTGGCTAGCGGTATTCGTGAGAGCGTGTCTAG 473

RESULT 10  
AA650659 527 bp mRNA linear EST 31-OCT-1997  
LOCUS 30773 Lambda-PRL2 Arabidopsis thaliana cdna clone 276G8T7, mRNA  
DEFINITION  
ACCESSION  
VERSION AA650659.1 GI:2580751  
KEYWORDS EST.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 527)  
Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H.,  
McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M.,  
Retzel, E. and Somerville, C.  
Genes Galore: a summary of methods for accessing results from  
large-scale partial sequencing of anonymous Arabidopsis cdna clones  
Plant Physiol. 106, 1241-1255 (1994)  
95148729  
MEDLINE  
PUBMED 7846151  
COMMENT  
Contact: Thomas Newman  
MSU-DOE Plant Research Laboratory  
Michigan State University  
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.  
Lansing, MI

Tel: 517-353-0854  
Fax: 517-353-9168  
Email: 22313cn@bm.cl.msu.edu  
Seq primer: T7 dye primer.

## FEATURES

Location/Qualifiers  
1..527  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/strain="var columbia"  
/db\_xref="taxon:3702"  
/clone="276G8T7"  
/clone\_lib="Lambda-PRL2"  
/note="Vector: lambda Zip-Lox; Site 1: Sal; Site 2: Not;  
quantities of 4 pools of mRNA. The mRNA derived from equal  
day germinated etiolated seedlings; 2) tissue culture  
grown roots; 3) staged plants half with 24 hour light  
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
same plants as 3 but aerial tissue (stems, flowers and  
siliques. The vector is BRU's lambda Zip-Lox. The cdna  
inserts were directionally cloned with Sal-Not arms using  
oligo dt primed cdna. "

## ORIGIN

Query Match 53.9%; Score 255.4; DB 9; Length 527;  
Best Local Similarity 80.2%; Pred. No. 7.3e-65;  
Matches 341; Conservative 0; Mismatches 21; Indels 63; Gaps 1;  
QY 50 CGAGGACTCTGCTACCAAGCGGAGGAAATCTGCGCGACTACTGAGGAGTCA 109  
DB 1 CGAGGAAGCAGCTCGCTACCAAGCGGAGGAAATCTGCAACGACTACCGGAGGAGTCA 60  
QY 110 AGAAACCTCACGTTACCGTCCCGGAACCGTCTTCGTGAGATTCTGTAATACCA 169  
DB 61 AGAAACCTCACGTTACCGTCTTCGTAACCGTCTTCGTGAGATTCTGTAATACCA 120  
QY 170 AGAGCACAGAGTGTGTGATCCGTAACTCTCTTTCAACGCTTGTTCGTGAAATCGCTC 229  
DB 121 AGAGCACAGAGTGTGTGATCCGTAACTCTTCATCCAGCTCTTGTTCGTGAAATCGCTC 180  
QY 230 AGATTACAAGACGGATCTGAGATCTCCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTG 289  
DB 181 AAGATTTCAA----- 190  
QY 290 CTGAACCATATTTGACGAGTCTGAGATCCAGAGCCATCGGTGTTAGCTCTTCAAGAG 349  
DB 191 -----GACGATCTGAGATCCAAAGCCATGCACTGTTAGCTCTTCAAGAG 237  
QY 350 CTGCGGAGCATATTTGGTGGGTTGTTTGAAGACACAAATCTGTGTGCCATTTCATGCAA 409  
DB 238 CTGCGGAGCATATTTGGTGGGTTGTTTGAAGACACAAATCTGTGTGCCATTTCATGCAA 297  
QY 410 AGAGGTTACGATATATGCTTAAGATGTTCAATGGCAAGAGGATTCGTGGAGAGCTG 469  
DB 298 AGAGGTTACGATATATGCTTAAGATGTTCAATGGCAAGAGGATTCGTGGAGAGCTG 357  
QY 470 CTTAG 474  
DB 358 CTTAG 362

RESULT 11  
BO791478

LOCUS BO791478 545 bp mRNA linear EST 30-JUL-2002  
DEFINITION E4401 Chinese cabbage etiolated seedling library Brassica rapa  
subsp. pekinensis cdna clone E4401, mRNA sequence.

ACCESSION  
VERSION BO791478  
KEYWORDS  
SOURCE  
ORGANISM

Brassica rapa subsp. pekinensis  
Brassica rapa subsp. pekinensis  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
 1 (bases 1 to 545)  
 Ryu,S.H., Yang,K.A., Lee,S.Y., Kim,H.-I., Cho,M.J. and Lim,C.O.  
 Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA  
 (2002)  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Lim, C.O.  
 Plant Molecular Biology & Biotechnology Research Centre  
 Gyeongsang National University  
 #900 Gazwa-dong, Jinju 660-701, Korea  
 Tel: 82 55 751 6255  
 Fax: 82 55 759 9363  
 Email: colim@nongae.gsnu.ac.kr  
 Seq primer: T7  
 FEATURES Location/Qualifiers  
 1..545  
 /organism="Brassica rapa subsp. pekinensis"  
 /mol\_type="mRNA"  
 /cultivar="Jangwon"  
 /sub\_species="pekinensis"  
 /db\_xref="taxon:51351"  
 /clone="E4401"  
 /tissue\_type="Etiolated seedling"  
 /lab\_host="XL-1 Blue"  
 /clone\_lib="Chinese cabbage etiolated seedling library"  
 /note="Vector: pSPORT 1; Site\_1: Sal I; Site\_2: Not I"

ORIGIN  
 Query Match 50.2%; Score 237.8; DB 13; Length 545;  
 Best Local Similarity 73.6%; Pred. No. 1.3e-59;  
 Matches 348; Conservative 0; Mismatches 62; Indels 63; Gaps 1;  
 QY 1 ATGGCTCTGACGAACAACTCAGAAATCACACGGAGGAAGAGCTCCGAGGACTCTG 60  
 Db |||||  
 QY 34 ATGGCTCTGACGAAGCAACCCCTGAAAGTCCACGGAGGCAAGACTCCGAGGAGCAG 93  
 Db |||||  
 QY 61 CTCGCTACCAAGCGCGGAGGAAATCTGCGCGACTACTTGGAGGAGTCAAGAAACCTCAC 120  
 Db |||||  
 QY 94 CTCGCCACCAAGCGCGGAGGAACTCTGCTCCGACCACCGGAGGAGTCAAGAAACCTCAC 153  
 QY 121 CGTTACCGTCCCGGACCGTCTCTGAGATTCGTAATACAGAAAGCAGACAG 180  
 Db |||||  
 QY 154 CGTTACCGTCCCGGACCGTGGCGCTCTGAGATCAGGAAGTACCAGAGAGCGCGAG 213  
 QY 181 TTGTTGATCCGTAACCTCTCTTCAACCTTGTTCGTAATCGCTCAAGATTACAG 240  
 Db |||||  
 QY 214 CTTTGAUCCGTAAGCTTCCCTTCAGCGTCTGGTTCGTAATCGCTCAGGACTCAG 273  
 QY 241 ACGGATCTGAGATTCAGAGCCATCGGTTAGCTCTTCAAGAAAGCTCTGAAGCATAT 300  
 Db |||||  
 QY 274 ACGGATCTGAGTTCCAGAGCCACGGGTGTTGGCTCTCAGGAGCGGCT----- 324  
 QY 301 TTGACGGATCTGAGATTCAGAGCCATCGGTTAGCTCTTCAAGAAAGCTCGGAGGCA 360  
 Db |||||  
 QY 325 -----GAGGCG 330  
 QY 361 TATTTGGTGGTTTGTGTAAGACACAAATCTGTGTGCCATTATGCAAGAGGGTTACG 420  
 Db |||||  
 QY 331 TACCTTGGTGGTTTGTGAGGATCTAATCTCTGCGGATTCACGCGAGAGGGTGACG 390  
 QY 421 ATAATGCCCTAAAGATTTCAATTTGGCAAGAGGATTCGTGGAGCGTCTTGA 473  
 Db |||||  
 QY 391 ATCATGCCCAAGGATATTCAGTTGGGAGGAGGATTCGTGGTGAACGTGCTTA 443

RESULT 12  
 AV552968  
 LOCUS AV552968 Arabidopsis thaliana roots Columbia Arabidopsis thaliana  
 DEFINITION cDNA clone RZ50d08R 5', mRNA sequence.  
 ACCESSION AV552968  
 VERSION AV552968.1 GI:8724381  
 KEYWORDS EST.

SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 545)  
 AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.  
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
 of 12,028 non-redundant expressed sequence tags from normalized and  
 size-selected cDNA libraries  
 JOURNAL DNA Res. 7 (3), 175-180 (2000)  
 MEDLINE 20363093  
 PUBMED 10907847  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.  
 FEATURES Location/Qualifiers  
 1..545  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="RZ50d08R"  
 /tissue\_type="roots"  
 /clone\_lib="Arabidopsis thaliana roots Columbia"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
 XhoI"

ORIGIN  
 Query Match 49.2%; Score 233; DB 9; Length 545;  
 Best Local Similarity 72.9%; Pred. No. 3.4e-58;  
 Matches 345; Conservative 0; Mismatches 63; Indels 63; Gaps 1;  
 QY 1 ATGGCTCTGACGAACAACTCAGAAATCACACGGAGGAAGAGCTCCGAGGACTCTG 60  
 Db 26 ATGGCTCTGACGAACAAACCGCTAGGAAATCCACCGGAGGTAAAGTCCCGAGGAGCAG 85  
 QY 61 CTCGCTACCAAGCGCGGAGGAAATCTGCGCGACTACTTGGAGGAGTCAAGAAACCTCAC 120  
 Db 86 CTTGCCACCAAGCGCGGAGGAAATCAGACCAACCGGAGGAGTCAAGAAACCTCAC 145  
 QY 121 CTTACCGTCCCGGAAACCGTCTCTCGTGAATTCGTAATACCAAGAGCACAG 180  
 Db 146 CGTTACCGTCCAGGAACCGTCTCTCTGAGATTCGTAGTACAGAGAGTACTGA 205  
 QY 181 TTGTTGATCCGTAACCTCTCTTTCAACGCTCTTGTGTGAAATCGCTCAAGATTACAG 240  
 Db 206 TTGTTGATCCGCAAGCTTCCCTTCCAGCGTCTCTGTTCCGGAATAGCTCAGGACTTCAAG 265  
 QY 241 ACGGATCTGAGATTCAGAGCCATCGGTTAGCTCTTCAAGAGCTCTGAAGCATAT 300  
 Db 266 ACGGATCTGAGTTTCAGAGTCACCGGTGTTAGCTCTTCAAGAGCGGCT----- 316  
 QY 301 TTGACGGATCTGAGATTCAGAGCCATCGGTTAGCTCTTCAAGAAAGCTCGGAGGCA 360  
 Db 317 -----GAGGCA 322  
 QY 361 TATTTGGTGGTTTGTGTAAGACACAAATCTGTGTGCCATTATGCAAGAGGGTTACG 420  
 Db 323 TACCTAGTTGGACTTTTCGAGGATCTAATCTCTCGCTATTTCAGCCCAAGAGGGTTACG 382  
 QY 421 ATAATGCCCTAAAGATTTCAATTTGGCAAGAGGATTCGTGGAGCGTCTTGA 473  
 Db 383 ATCATGCCCTAAGGACATTCATTTGCTAGAGATTCGTGGAGAACGTCCTTA 435

RESULT 13  
 CD830234  
 LOCUS CD830234 Arabidopsis thaliana roots Columbia Arabidopsis thaliana  
 DEFINITION cDNA clone RZ50d08R 5', mRNA sequence.  
 ACCESSION CD830234  
 VERSION BN40.044M18F011221 BN40 Brassica napus cDNA clone BN40044M18, mRNA  
 KEYWORDS sequence.



```

ACCESSION   CD830234
VERSION     CD830234.1  GI:32512174
KEYWORDS    EST.
SOURCE      Brassica napus (rape)
ORGANISM    Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 603)
AUTHORS     Genoplante.
            Genoplante, a major partnership french program in plant genomics
            Unpublished (2003)
TITLE       Genoplante
JOURNAL
COMMENT     Contact: Genoplante
            Location/Qualifiers
FEATURES             source
             1..603
                /organism="Brassica napus"
                /mol_type="mRNA"
                /cultivar="Jet neuf"
                /db_xref="taxon:3708"
                /clone="BN4004M18"
                /tissue_type="seed"
                /clone_lib="BN40"

ORIGIN
Query Match      49.2%; Score 233; DB 14; Length 603;
Best Local Similarity 72.9%; Pred. No. 3.6e-58;
Matches 345; Conservative 0; Mismatches 65; Indels 63; Gaps 1;

Qy  1  ATGGCTCGTAGAGCAAACTGCAAGAAATACACAGGAGGAAAGCTCCGAGGACTCTG 60
Db  69  ATGGCTCGTAGAGCAAAACCGCTAGAAAGTCCACGAGGAGCAAGTCCGAGGAGCAG 128
Qy  61  CTCGCTACCAAGGCGGAGGAAATCTGCGCGGACTACTGAGGAGGTCAAGAAACCTTAC 120
Db  129  CTCGCCACCAAGGCGGAGGAAATCTGCTCCGACGACCGGAGGAGTCAAGAGCCGCAC 188
Qy  121  CGTTACCGTCCGAGAACCGTCCGCTCTGCTGAGATTCTTAATACCAAGAGACACAGAG 180
Db  189  CGTTACCGTCCGAGAACCGTCCGCTCTGCTGAGATCAGGAAGTACCAGAGACCGCAG 248
Qy  181  TTGTTGATCCGTAAACTTCCTTTTCAAGCTCTTGTTCGTAATCGCTCAAGATTACAAG 240
Db  249  CTCCTTGATCCGTAGCTTCTTTTCCAGCGTCTGGTTCGTGAATCGCTCAGGACTTCAA- 307
Qy  241  ACGGATCTGAGATTCAGAGCCATGCGGTGTAGTCTCTTCAAGAGCTGCTGAAGCATAT 300
Db  308  ----- 307
Qy  301  TTGACGGATCTGAGATTCAGAGCCATGCGGTGTAGTCTCTTCAAGAGCTGCGAGGCA 360
Db  308  --GACGATCTGAGATTCAGAGCCATGCGGTGTAGTCTCTTCAAGAGCTGCGAGGCG 365
Qy  361  TATTGTTGGGTTTGTGTTGAAGACACAAATCTGTGTGCCATTCATGCAAGAGGTTACG 420
Db  366  TACCTCGTGGGTTTGTTCGAGGACTAATCTCTGCGGATTCAGCGAGAGGCTCAGC 425
Qy  421  ATATGCTTAAGATGTTCAATTGCAAGAGGATTCGTGAGAGCGTGTCTTA 473
Db  426  ATCATGCCCAAGGATATTTCAGTTGGCCAGGAGGATTCGTGGTGAACGTGCTTA 478

RESULT 14
AI998977/c 611 bp mRNA linear EST 08-SEP-1999
LOCUS      701547802 A. thaliana, Columbia Col-0, rosette-2 Arabidopsis
DEFINITION thaliana cDNA clone 701547802, mRNA sequence.

```

```

ACCESSION   AI998977
VERSION     AI998977.1  GI:5845882
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 611)
AUTHORS     Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,
            Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P.,
            Gorsone, G., Burns, D., Griffin, J., Mouanoutou, M., Nguyen, D.,
            Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C.,
            Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., and
            Nobrega, A., Murry, L., Turner, C., Krikorian, S., Elder, L., and
            Hanson, D.
            Arabidopsis thaliana Gene Expression MicroArray
            Unpublished (1999)
TITLE       Arabidopsis thaliana
JOURNAL
COMMENT     Contact: David Smoller, Ph.D.
            Genome Systems, Inc., a wholly owned subsidiary of Incyte
            Pharmaceuticals, Inc.
            4633 World Parkway Circle, St. Louis, MO 63134, USA
            Tel: 877-577-2733
            Fax: 314-427-3324
            Email: service@genomesystems.com.
            Location/Qualifiers
FEATURES             source
             1..611
                /organism="Arabidopsis thaliana"
                /mol_type="mRNA"
                /cultivar="Columbia Col-0"
                /db_xref="taxon:3702"
                /clone="701547802"
                /tissue_type="rosette"
                /dev_stage="4 - 7 weeks"
                /clone_lib="A. thaliana, Columbia Col-0, rosette-2"
                /notes="Vector: pSPORT; Site 1: NotI; Site 2: SalI; cDNA
                library was derived from untreated rosette tissue from
                Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks.
                Plants were grown in 1:1:1 peat moss/vermiculite/perlite
                soil at 22 deg. C +/- 3 deg. C under constant light, and
                watered with fertilizer. cDNA synthesis was initiated
                using a NotI-oligo(dT) primer. Double-stranded cDNA was
                blunted, ligated to SalI adaptors, digested with NotI,
                size-selected, and cloned into the NotI and SalI sites of
                the pSPORT vector."

ORIGIN
Query Match      49.2%; Score 233; DB 9; Length 611;
Best Local Similarity 72.9%; Pred. No. 3.6e-59;
Matches 345; Conservative 0; Mismatches 65; Indels 63; Gaps 1;

Qy  1  ATGGCTCGTAGAGCAAACTGCAAGAAATACACAGGAGGAAAGCTCCGAGGACTCTG 60
Db  547  ATGGCTCGTAGAGCAAAACCGCTAGGAAATCCACGAGGAGTAAAGTCCCGAAGCAG 488
Qy  61  CTCGCTACCAAGGCGGAGGAAATCTGCGCGCACTACTCGAGGAGTCAAGAAACCTTCA 120
Db  487  CTTGCCACCAAGGCGGAGGAAATCAGCACCAACACCGGAGGAGTCAAGAGCCTTCA 428
Qy  121  CGTTACCGTCCGAGAACCGTCCGCTCTTCTGAGATTCTGTAATACCAAGAGACAGAG 180
Db  427  CGTTACCGTCCGAGAACCGTCCGCTCTTCTGAGATTCTGTAATACCAAGAGAGTACTGAA 368
Qy  181  TTGTTGATCCGTAAACTTCCTTTTCAAGCTCTTCTGTAATCGCTCAAGATTACAAG 240
Db  367  TTGTTGATCCGCAAGCTTCCCTTCCAGCGTCTCGTTCGGGAAATAGCTCAGGACTTCAAG 308
Qy  241  ACGGATCTGAGATTCAGAGCCATGCGGTGTAGTCTCTTCAAGAGCTGCTGAAGCATAT 300
Db  307  ACGGATCTGAGTTTCAGAGTTCAGCGGTGTAGTCTCTTCAAGAGGCGGCT----- 257
Qy  301  TTGACGATCTGAGATTCCAGAGCCATGCGGTGTAGTCTCTTCAAGAGCTGCGGAGGCA 360

```



```

Db      256 -----GAGGCA 251
Qy      361 TATTGGTGGGTTTGTGTAAGACACAAATCTGTGTGCCATTATGCAAGAGGGTTACG 420
Db      250 TACCTAGTTGGACTTTTCGAGGACTAAATCTCTGCGCTATTACGCCAAGAGGGTTACG 191
Qy      421 ATAATGCTTAAGATGTTCAATTGGCAGAAGAGGATTGCTGGAGAGCGTGCTTA 473
Db      190 ATCATGCCCTAAGGACATTCATTTGGCTAGAGAAATTCGTGGAGAACGTGCTTA 138

RESULT 15
CD822585
LOCUS   BN25.045M06F020108 BN25 Brassica napus cDNA clone BN25045M06, mRNA
DEFINITION
sequence.
ACCESSION   CD822585
VERSION     CD822585.1 GI:32504525
KEYWORDS    EST.
SOURCE      Brassica napus (rape)
ORGANISM    Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 643)
AUTHORS     Genoplante.
TITLE       Genoplante, a major partnership french program in plant genomics
JOURNAL     Unpublished (2003)
COMMENT     Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).

FEATURES             source
     1..643
     /organism="Brassica napus"
     /mol_type="mRNA"
     /cultivar="Jet neuf"
     /db_xref="taxon:3708"
     /clone="BN25045M06"
     /tissue_type="seed"
     /clone_lib="BN25"

ORIGIN
Query Match      49.2%; Score 233; DB 14; Length 643;
Best Local Similarity 72.9%; Pred. No. 3.7e-58;
Matches 345; Conservative 0; Mismatches 65; Indels 63; Gaps 1;

Qy      1 ATGGTCTGTACGACCAACTCGAAGAAATCACACGGAGGAAAGCTCCGAGGACTCTG 50
Db      69 ATGGCTGTGTACGAAACAAACCGCTAGAAAGTCCACGGAGGCAAAAGCTCCCGAAGCAG 128
Qy      61 CTCGTCTACCAAGCGCGGAGGAAATCTGCGCGACTACTGGAGGAGTCAAGAAACCTCAC 120
Db      129 CTCGCCACCAAGCGCGGAGGAAATCTGCTCGACGACCGGAGGAGTCAAGAGCGGCAC 188
Qy      121 CATTACCGTCCCGAAACCGTCGCTCTTCGTGAGATTCGTAATACAGAAAGACACAGAG 180
Db      189 CGTTACCGTCCAGGAACCGTCGCGCTTCGTGAGATCAGGAAGTACCAGAAAGACGACGGAG 248
Qy      181 TTGTTGATCCGTAACCTCTTTCAACGCTTGTTCTGTGTAATCGCTCAAGATTACAG 240
Db      249 CTCCTTGATCCGTAAGTCTCTTTCCAGCGTCTGTTTCGTAATCGCTCAGGACTTCAA- 307
Qy      241 ACGGATCTGAGATTCACAGGCCATGCGGTGTAGTCTTTCAAGAAGCTGCTGAAGCATAT 300
Db      308 -----
Qy      301 TTGACGGATCTGAGATTCACAGACCGATGCGGTGTAGCTTTCAAGAACTCGGAGGCA 360

```

```

Db      308 --GACGGATCTGAGGTTCCAGAGCCACGCCGTGTGTGGCTCTCCAGGAGGCGCGAGGCG 365
Qy      361 TATTGGTGGGTTTGTGTAAGACACAAATCTGTGTGCCATTATGCAAGAGGGTTACG 420
Db      366 TACCTCGTGGGTTTGTTCGAGGATACTAAATCTCTGCGGATTCACGCCAAGAGGGTTACG 425
Qy      421 ATAATGCTTAAGATGTTCAATTGGCAGAAGAGGATTGCTGGAGAGCGTGCTTA 473
Db      426 ATCATGCCCAAGGATATTCAGTTGGCCAGGAGGATTCGTGGTGAACGTGCTTA 478

```

Search completed: June 28, 2004, 18:01:36  
Job time : 1648.18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 18:04:41 ; Search time 43 seconds  
(without alignments)  
364.633 Million cell updates/sec

Title: AAC50492

Perfect score: 852

Sequence: 1 MADBEKLPFGWEKMRSSSG.....GEMSGPVFTDGIHILRTE 163

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_78:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	852	100.0	163	2 S68520	Pin1 protein - hum
2	814	95.5	165	2 JC7136	peptidylprolyl iso
3	456.5	53.6	166	2 T08426	Pin1 protein.homol
4	388.5	45.6	175	2 T41093	peptidyl-prolyl ci
5	356.5	41.8	170	2 S52764	ESSI protein - yea
6	301	35.3	119	2 E45559	probable peptidyl-
7	180.5	21.2	292	2 S15269	post-translocation
8	177	20.8	247	2 A96934	peptidyl-prolyl ci
9	158	18.5	142	2 E86392	hypothetical prote
10	157.5	18.5	126	2 T31601	hypothetical prote
11	153.5	18.0	331	2 AF3267	peptidyl-prolyl ci
12	149	17.5	273	2 S52412	major antigenic pe
13	148.5	17.4	333	2 A83797	protein secretion
14	147	17.3	655	2 E82712	peptidyl-prolyl ci
15	145	17.0	347	2 A81794	probable rotamase
16	145	17.0	348	2 B81216	peptidyl-prolyl ci
17	138.5	16.3	621	2 A83421	peptidyl-prolyl ci
18	138	16.2	93	2 E91217	peptidylprolyl iso
19	138	16.2	93	2 F85063	peptidylprolyl iso
20	138	16.2	93	2 S48658	peptidylprolyl iso
21	137	16.1	93	2 AE0923	peptidyl-prolyl ci
22	127.5	15.0	320	2 H09970	hypothetical prote
23	127	14.9	93	2 AH0471	peptidylprolyl iso
24	126.5	14.8	92	2 F83395	peptidyl-prolyl ci
25	124	14.6	300	2 AF2989	peptidyl-prolyl ci
26	124	14.6	315	2 C98294	cell binding facto
27	120.5	14.1	92	2 AD8296	peptidyl-prolyl ci
28	119.5	14.0	434	2 AD0061	peptidylprolyl iso
29	117	13.7	293	2 AC1352	post-translocation

30	116	13.6	428	2 B30636	survival protein [
31	116	13.6	428	2 B85487	survival protein {
32	116	13.6	428	2 B84726	probable peptidylp
33	114.5	13.4	431	2 D82323	survival protein S
34	114.5	13.4	708	2 I83196	NEDD-4 ORF - mouse
35	114	13.4	93	2 C83123	peptidyl-prolyl ci
36	114	13.4	291	2 AF1722	post-translocation
37	112.5	13.2	266	2 S02510	nifM protein - Kle
38	112	13.1	820	2 T46412	ubiquitin-protein
39	110.5	13.0	293	2 H43706	nifM protein - Azo
40	109.5	12.9	221	2 T52622	probable peptidylp
41	109	12.8	428	2 AH0513	survival protein S
42	108.5	12.7	353	2 D97295	parvulin-like pept
43	108	12.7	336	2 D70113	basic membrane pro
44	107.5	12.6	120	2 T26779	hypothetical prote
45	107.5	12.6	887	2 S70642	ubiquitin ligase N

#### ALIGNMENTS

##### RESULT 1

S68520

Pin1 protein - human

C:Species: Homo sapiens (man)

C>Date: 15-Nov-1996 #sequence\_revision 27-Feb-1997 #text\_change 29-Sep-1999

C:Accession: S68520

R:Ida, K.F.; Hanes, S.D.; Hunter, T.

Nature 380, 544-547, 1996

A:Title: A human peptidyl-prolyl isomerase essential for regulation of mitosis.

A:Reference number: S68520; MUID:96195064; PMID:8606777

A:Accession: S68520

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-163 <LUK>

A:Cross-references: EMBL:U49070; NID:g1332709; PIDN:AAC50492.1; PID:g1332710

C:Genetics:

A:Gene: GDB:PIN1; dod

A:Cross-references: GDB:5218381

A:Superfamily: Yeast ESSI protein; WW repeat homology

F:5-43/Domain: WW repeat homology <WW1>

Query Match	100.0%	Score 852;	DB 2;	Length 163;
Best Local Similarity	100.0%;	Pred. No. 9.4e-72;		
Matches 163;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MADBEKLPFGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPARVRCSHL	60	
Db	1	MADBEKLPFGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPARVRCSHL	60	
QY	61	LVKHSQRRPSSWROEKITRKEALELINGYIOKIKSGEEDFESLASQFSDCSSAKRG	120	
Db	61	LVKHSQRRPSSWROEKITRKEALELINGYIOKIKSGEEDFESLASQFSDCSSAKRG	120	
QY	121	DLGAFSRQMQKPPEDASFALRTGEMSGPVFTDGIHILRTE	163	
Db	121	DLGAFSRQMQKPPEDASFALRTGEMSGPVFTDGIHILRTE	163	

##### RESULT 2

JC7136

peptidylprolyl isomerase (EC 5.2.1.8) - mouse

N:Alternate names: parvulins, pin1 protein

C:Species: Mus musculus (house mouse)

C>Date: 04-Mar-2000 #sequence\_revision 04-Mar-2000 #text\_change 18-Aug-2000

C:Accession: JC7136

R:Fujimori, F.; Takahashi, K.; Uchida, C.; Uchida, T.

Biochem. Biophys. Res. Commun. 265, 658-663, 1999

A:Title: Mice lacking Pin1 develop normally, but are defective in entering cell cycle f:

A:Reference number: JC7136; MUID:20070807; PMID:10600477

A:Accession: JC7136

A:Molecule type: mRNA

A:Residues: 1-165 <FUJ>

A;Cross-references: DBJ:AB09691; NID:g6468199; PIDN:BAA7037.1; PID:g6468200  
A;Comment: This protein is a member of peptidyl prolyl cis/trans isomerase family. It is  
tes in oligopeptides, controls G2/M transition of HeLa cells, regulates entry into G1 an  
C;Genetics:  
A;Gene: Pin1  
C;Superfamily: yeast ESS1 protein; WW repeat homology  
C;Keywords: cis-trans-isomerase; mitosis  
F;5-43/Domain: WW repeat homology <WWR>

Query Match 95.5%; Score 814; DB 2; Length 165;  
Best Local Similarity 95.2%; Pred. No. 3.2e-68;  
Matches 157; Conservative 3; Mismatches 3; Indels 2; Gaps 1;  
QY 1 MADEKLPQGWKMRSSGRVYFNHITNASQWPPSGSGG--KNGQGEPAVRCS 58  
DB 1 MADEKLPQGWKMRSSGRVYFNHITNASQWPPSGSTVGGSSKNGQGEPAVRCS 60  
QY 59 HLLVKSQSRPPSSWQEKITRKEALELINGYIQIKSGEDFSLASQFSDCSAKA 118  
DB 61 HLLVKSQSRPPSSWQEKITRKEALELINGYIQIKSGEDFSLASQFSDCSAKA 120  
QY 119 RGDLPFSGQMQKPFEDAFALRTGEMSGPVFTDSCIHIILTE 163  
DB 121 RGDLPFSGQMQKPFEDAFALRTGEMSGPVFTDSCIHIILTE 165

RESULT 3  
T08426  
Pin1 protein homolog dodo - fruit fly (Drosophila melanogaster)  
C;Species: Drosophila melanogaster  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 17-Nov-2000  
C;Accession: T08426  
R;Maleszka, R.; de Couet, H.G.; Miklos, G.L.  
Proc. Natl. Acad. Sci. U.S.A. 95, 3731-3736, 1998  
A;Title: Data transferability from model organisms to human beings: insights from the fu  
A;Reference number: Z16415; MUID:98188272; PMID:9520435  
A;Accession: T08426  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-166 <MAL>  
A;Cross-references: EMBL:AF01777; NID:g3004652; PIDN:AAC28408.1; PID:g3004661  
A;Experimental source: strain Cantons  
C;Genetics:  
A;Gene: dod  
A;Cross-references: FlyBase:FBgn0024251  
A;Introns: 20/1, 131/1  
C;Superfamily: yeast ESS1 protein; WW repeat homology  
F;5-43/Domain: WW repeat homology <WWR>

Query Match 53.6%; Score 456.5; DB 2; Length 166;  
Best Local Similarity 56.7%; Pred. No. 4.5e-35;  
Matches 93; Conservative 16; Mismatches 52; Indels 3; Gaps 2;  
QY 1 MADEKLPQGWKMRSSGRVYFNHITNASQWPPSS--GNSSSGKNGQGE-PAVRCS 57  
DB 1 MPDAEQLPDQGWKMRSSGRVYFNHITNASQWPPSSQDQTEPAKTKGGSGAGGDPDVRHC 60  
QY 58 HLLVKSQSRPPSSWQEKITRKEALELINGYIQIKSGEDFSLASQFSDCSAK 117  
DB 61 LHLVKSQSRPPSSWQEKITRKEALELINGYIQIKSGEDFSLASQFSDCSAK 120  
QY 118 ARGDLGAFSRQMQKPFEDAFALRTGEMSGPVFTDSCIHIILR 161  
DB 121 RGDLPFSGQMQKPFEDAFALRTGEMSGPVFTDSCIHIILR 164

RESULT 4  
T41093  
peptidyl-prolyl cis-trans isomerase - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Aug-2000  
C;Accession: T41093  
R;Furnelle, B.; Goffeau, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, September 1998  
A;Reference number: Z21964  
A;Accession: T41093  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-175 <PUR>  
A;Cross-references: EMBL:AL031535; PIDN:CAA20742.1; GSPDB:GN00068; SPDB:SPCC16C4.03  
A;Experimental source: strain 972h; cosmid cl6C4  
C;Genetics:  
A;Gene: SPDB:SPCC16C4.03  
A;Map position: 3  
A;Introns: 134/2  
C;Superfamily: yeast ESS1 protein; WW repeat homology  
F;4-42/Domain: WW repeat homology <WWR>

Query Match 45.6%; Score 388.5; DB 2; Length 175;  
Best Local Similarity 48.5%; Pred. No. 9.8e-29;  
Matches 82; Conservative 21; Mismatches 51; Indels 15; Gaps 2;  
QY 7 LPPGWEKMRSSGRVYFNHITNASQWPPSGSGGK-----NGQGE 52  
DB 6 LPKPMIVKISRPNRPFFNTTHESLNEPPAATDMAALKKFIANELQESVTPTEASNP 65  
QY 53 ARVRCSHLLVKSQSRPPSSWQEKITRKEALELINGYIQIKSGEDFSLASQFSD 112  
DB 66 -KIRASHLLVKSQSRPPSSWQEKITRKEALELINGYIQIKSGEDFSLASQFSD 124  
QY 113 CSAXARGDLGAFSRQMQKPFEDAFALRTGEMSGPVFTDSCIHIILR 161  
DB 125 CSAXARGDLGAFSRQMQKPFEDAFALRTGEMSGPVFTDSCIHIILR 173

RESULT 5  
S52764  
ESS1 protein - yeast (Saccharomyces cerevisiae)  
N;Alternate names: protein J1452; protein YUR017c; Ptf1 protein  
C;Species: Saccharomyces cerevisiae  
C;Date: 19-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 29-Sep-1999  
A;Accession: S52764; S5207; S57032; S07867; S65929  
R;Hani, J.; Stumpf, G.; Domdey, H.  
submitted to the EMBL Data Library, March 1995  
A;Description: Ptf1 encodes an essential protein in Saccharomyces cerevisiae, which show  
A;Reference number: S52762  
A;Accession: S52764  
A;Molecule type: DNA  
A;Residues: 1-170 <HAN>  
A;Cross-references: EMBL:X85972; NID:g758283; PIDN:CAA59961.1; PID:g758286  
R;de Haan, M.; Smits, P.H.M.; Grivell, L.A.  
submitted to the EMBL Data Library, May 1995  
A;Reference number: S55183  
A;Accession: S55207  
A;Molecule type: DNA  
A;Residues: 'MLIVLIPTNRLRHNDAP', 1-170 <DEH>  
A;Cross-references: EMBL:X87611; NID:g854567; PIDN:CAA60941.1; PID:g854592  
R;de Haan, M.; Grivell, L.A.; Smits, P.H.M.  
submitted to the Protein Sequence Database, September 1995  
A;Reference number: S56771  
A;Accession: S57032  
A;Molecule type: DNA  
A;Residues: 'MLIVLIPTNRLRHNDAP', 1-170 <ZAG>  
A;Cross-references: EMBL:Z49517; NID:g1015651; PIDN:CAA89541.1; PID:g1015652; MIPS:YUR01  
R;Hanes, S.D.; Shank, P.R.; Bostian, K.A.  
Yeast 5, 55-72, 1989  
A;Title: Sequence and mutational analysis of ESS1, a gene essential for growth in Saccha  
A;Reference number: S07867; MUID:89189095; PMID:2648698  
A;Accession: S07867  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 'MLIVLIPTNRLRHNDAP', 1-7, 'S', '9-16, 'A', '18-126, 'AASA', '133, 'S', '135-136, 'ARC',  
A;Experimental source: strain DBY864  
R;Hani, J.; Stumpf, G.; Domdey, H.  
FEBS Lett. 365, 198-202, 1995  
A;Title: Ptf1 encodes an essential protein in Saccharomyces cerevisiae, which shows stro





E82712  
pepidyl-prolyl cis-trans isomerase XF1191 [imported] - Xylella fastidiosa (strain 9A5C)  
C Species: Xylella fastidiosa  
C Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C Accession: E82712  
R Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
Nature 406, 151-157, 2000  
A Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A Reference number: A8515; MUID:20355717; PMID:10910347  
A Note: for a complete list of authors see reference number A59328 below  
A Accession: E82712  
A Status: preliminary  
A Molecule type: DNA  
A Residues: 1-655 <STM>  
A Cross-references: GB:AEO03953; GB:AEO03849; NID:g9106156; PIDN:AAF84001.1; GSPDB:GN00.  
A Experimental source: strain 9A5C  
F Simpson, A.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; ; Briones, M.R.S.; Bueno, M.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, I. as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohr, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krisger, J.E.; Kuramae, E.E.; Laig, chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, I.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y., Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasal, A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; ;  
A Reference number: A59328  
A Contents: annotation  
C Genetics:  
A Gene: XF1191

Query Match 17.3%; Score 147; DB 2; Length 655;  
Best Local Similarity 28.5%; Pred. No. 1.2e-05;  
Matches 39; Conservative 23; Mismatches 57; Indels 18; Gaps 3;

Qy 30 NASOWEPGSGNSS-----GGKNGQGEPAVRVCSHLLVKHSQRSSWRQEKITRKE 83  
Db 257 DAHLPRPTTSEAVLRKYRYDAQORGEAQNRKANHLITAGADAASQKVAAKAADVE 316  
Qy 84 EAELINGYIQIKSGEDEFESLAS-QFSCSSAKARGDLGAFSRGOMQRPEDASFALR 142  
Db 317 EA-----RKPGVDFAALARINSQDPGSDAGGLGWQGRGMWVPFDVLFAMK 365  
Qy 143 TGEMSGPVFTDSGHII 159  
Db 366 VGEVVGVPIKTEFGNHVI 382

RESULT 15  
A81794  
Probable rotamase NMA2206 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)  
C Species: Neisseria meningitidis  
C Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C Accession: A81794  
R Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Moret, R.; Parkhill, S.; Jacobs, K.; Leather, S.; Moule, S.; Muncall, M.A.; Rajandream, A.; Holtz, J.

A;Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.  
 A;Reference number: A81775; MUID:20222556; PMID:10761919  
 A;Accession: A81794  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-347 <PAR>  
 A;Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85417.1; PID:g73808.  
 A;Experimental source: serogroup A, strain Z2491  
 C;Genetics:  
 A;Gene: NMA2206

Query Match 17.0%; Score 145; DB 2; Length 347;  
Best Local Similarity 37.1%; Pred. No. 8.3e-06;  
Matches 43; Conservative 15; Mismatches 44. Indels

Qy	49	QCEPAR-VRCSHLLVKHSOSRRPSSWRQEKITRTKEEALELINGYIQIKSGEEDFESLA	107
Db	201	EGAPLQYRAQHILIK-ADSENAAVGAESTIRKIYGEA-----RSG-TDFSSLA	247
Qy	108	SQFSDCSSAKARGDLGAFSRGQMKPFEDASPALRTGEMSGPVFTDSGTHILRTE	163
Db	248	ROYSQDASAGNGDGLGWFADGVMPAFEEAVHALKPGQVGAPVRTQFGWHIIKLE	303

Search completed: June 28, 2004, 18:39:59  
Job time : 45 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 16:36:51 ; Search time 26 Seconds  
(without alignments)  
326.440 Million cell updates/sec

Title: AAC50492

Perfect score: 852

Sequence: 1 MADEKLPFGWEKMRSSG.....GEMSGPVFTSDGIHILRTE 163

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	852	100.0	163	1 PIN1 HUMAN	Q13526 homo sapien
2	814	95.5	165	1 PIN1 MOUSE	Q9qur7 mus musculus
3	456.5	53.6	166	1 DOD DROME	P54353 drosophila
4	428.5	50.3	182	1 SSF1 NEUCR	O60045 neurospora
5	388.5	45.6	175	1 PIN1 SCHPO	O74448 schizosacch
6	356.5	41.8	190	1 ESS1 YEAST	P22696 saccharomyc
7	301	35.3	119	1 PIN1 ARATH	Q96142 arabidopsis
8	297	34.9	100	1 PIN1 HUMAN	Q15428 homo sapien
9	296.5	34.8	121	1 PIN1 MALDO	Q94500 malus domes
10	283	33.2	118	1 PIN1 DIGLA	Q91ek8 digitalis l
11	180.5	21.2	292	1 PRSA_BACSU	P24327 bacillus su
12	152	17.8	131	1 PIN4 HUMAN	Q9Y237 homo sapien
13	152	17.8	131	1 PIN4 MOUSE	Q9CWW6 mus musculus
14	149	17.5	273	1 CBP2_CAMJE	Q46105 campylobact
15	138	16.2	92	1 PPTC_ECOLI	P39159 escherichia
16	137	16.1	92	1 PPTC_SALTY	Q916s3 salmonella
17	118	13.8	421	1 GAS7 MOUSE	Q60780 mus musculus
18	118	13.8	422	1 GAS7 RAT	O55148 rattus norv
19	116	13.6	428	1 SURA_ECOLI	P21202 escherichia
20	116	13.6	748	1 SUP2 HUMAN	Q9ha04 homo sapien
21	115.5	13.6	864	1 ITCH MOUSE	Q8c863 mus musculus
22	115.5	13.6	903	1 ITCH HUMAN	Q96j02 homo sapien
23	115	13.5	412	1 GAS7 HUMAN	O60861 homo sapien
24	114.5	13.4	887	1 NED4 MOUSE	P46935 mus musculus
25	114	13.4	870	1 WNP2 HUMAN	O00308 homo sapien
26	114	13.4	870	1 WNP2 KLEPN	Q9dbh0 mus musculus
27	112.5	13.2	266	1 NIFM_KLEPN	P08534 klebsiella
28	110.5	13.0	293	1 NIFM_AZOCH	P23119 azotobacter
29	109	12.8	518	1 WNP1 MOUSE	Q8bz23 mus musculus
30	108.5	12.7	922	1 WNP1 HUMAN	Q9h0m0 homo sapien
31	107.5	12.6	887	1 NED4 RAT	Q62940 rattus norv
32	107	12.6	299	1 Y175_HELPJ	Q9zmq7 helicobacte
33	107	12.6	299	1 Y175_HELPY	P56112 helicobacte

## RESULT 1

PIN1 HUMAN STANDARD; PRT; 163 AA.

AC Q13526;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, last sequence update)

DT 10-OCT-2003 (Rel. 42, last annotation update)

DE Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (BC 5.2.1.8)

DE (Rotamase Pin1) (PPIase Pin1).

GN PIN1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_taxID=9606;

RN [1]

SEQUENCE FROM N.A.

RX MEDLINE=96195064; PubMed=8606777;

RA Lu K.P., Hanes S.D., Hunter T.;

RT "A human peptidyl-prolyl isomerase essential for regulation of

mitosis."

RL Nature 380:544-547(1996).

RN [2]

SEQUENCE FROM N.A.

RP TISSUE=Lung;

MEDLINE=22388257; PubMed=12477932;

RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

Fahy J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,

Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalius D.E.,

Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;

"Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences."

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [3]

X-RAY CRYSTALLOGRAPHY (1.35 ÅNGSTRÖMS).

MEDLINE=97344079; PubMed=9200606;

RA Ranganathan R., Lu K.P., Hunter T., Noel J.P.;

"Structural and functional analysis of the mitotic rotamase Pin1

suggests substrate recognition is phosphorylation dependent."

RT

Cell 89:875-886(1997).

CC -!- FUNCTION: Essential PPIase that regulates mitosis presumably by

interacting with NIMA and attenuating its mitosis-promoting

activity. Displays a preference for an acidic residue N-terminal

to the isomerized proline bond. Catalyzing pSer/Thr-Pro cis/trans

CC







Best Local Similarity 56.7%; Pred. No. 2e-35;  
Matches 93; Conservative 16; Mismatches 52; Indels 3; Gaps 2;

```
QY 1 MADEKLPGEKMSRSGRVYFNHTNASHQWERS--GNSSGGKNGQGE-PARVRC 57
Db 1 MPDAQLPDGKEKTSRSTGMSYLYNMTYKESQMDQPTPAKAGGSGAGGDADEVHC 60
QY 58 SHLLVKHSQSRPPSRWROEKIRTKKEALELINGYQIKKSGEEDFESLASQFDCSSAK 117
Db 61 LHLVLKHSQSRPPSRWRANIRTKKEAQLLEVTNRKLVQEOATFDELARSYDCSSAK 120
QY 118 ARGDLGAFSRGQKPPEDAFALRTGEMSGEPVFTDGIHILR 161
Db 121 RGGDLGKFGKQMAAFEDAAFKLVNQLSGIVDSGLHILR 164
```

## RESULT 4

```
SSPI_NEUCR
ID SSPI_NEUCR STANDARD; PRT; 182 AA.
AC O60045;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase sspl (EC 5.2.1.8) (PPIase sspl).
GN SSPI.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=99041963; PubMed=9822668;
RA Kops O., Eckerskorn C., Hottenrott S., Fischer G., Mi H.,
RA Tropea M.;
RT "Sspl, a site specific parvulin homolog from N.crassa active in
RL protein folding";
RL J. Biol. Chem. 273:31971-31976 (1998).
CC -!- FUNCTION: Site-specific PPIase with respect to the amino acid N-
CC terminal to the proline residue. Peptides with glutamate,
CC phosphoserine, or phosphothreonine in the -1 position are the best
CC substrates. It is not only able to isomerize small peptides but
CC is also active in protein folding.
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -!- SIMILARITY: Contains 1 WW domain.
CC -!- SIMILARITY: Belongs to the ppiC/parvulin rotamase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ004603; CAA06818.1; -
CC HSP: Q13526; IPIN.
CC InterPro: IPR000297; Rotamase.
CC InterPro: IPR001202; WW_Rsp5_WWP.
CC Pfam: PF00639; Rotamase; 1.
CC Pfam: PF00397; WW; 1.
CC SMART: SM00456; WW; 1.
CC PROSITE: PS50198; PPIase 2; 1.
CC PROSITE: PS01159; WW_DOMAIN_1; 1.
CC PROSITE: PS50020; WW_DOMAIN_2; 1.
CC Isonomase; Rotamase.
CC DOMAIN 7 41 WW
CC FT DOMAIN 60 71 GLN-RICH.
CC FT DOMAIN 71 182 PPI.
CC SEQUENCE 182 AA; 20673 MW; AC8D0497DFF7339 CRC64;
```

Query Match 50.3%; Score 428.5; DB 1; Length 182;  
Best Local Similarity 50.8%; Pred. No. 8.7e-33;

Matches 90; Conservative 20; Mismatches 50; Indels 17; Gaps 2;

```
QY 4 BEKLPPGWEKMSRSGRVYFNHTNASHQWERSG-----NSSSGKN 47
Db 6 ETGLPEDVEVHSQKLPFYFNSTATKTSRWEPPSGTDVVKLTMYAKYHSFTSQOQQQ 65
QY 48 GQGP-ARVRCSHLLVKHSQSRPPSRWROEKIRTKKEALELINGYQIKKSGEEDFESL 106
Db 66 QQQQPOGKIRCAHLLVKHSQSRPPSRWRESEITTKREALTTLOGFQRKSGSISIGEL 125
QY 107 ASQFSCSSAKARGDLGAFSRGQKPPEDAFALRTGEMSGPVFTDGIHILRTE 163
Db 126 ALTESDCSSARKRGDLGFGFGDMQKBEFEDAAFKLPEISDVDTASGLHLIERLE 182
```

## RESULT 5

```
PINI_SCHPO
ID PINI_SCHPO STANDARD; PRT; 175 AA.
AC O74448;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase pinl (EC 5.2.1.8).
GN PINI OR SPCL164.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetiales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=2184801; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Meeney P., Moule S., Mungall K., Murphy L., Niblett D., Odel C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Welljens I., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Gabel C., Fuchs M., Fritz C., Rieger M., Schaefer M., Mueller-Auer S.,
RA Borzym K., Langer I., Beck A., Leinhardt H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Coffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucabart M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880 (2002).
RN [2]
RP FUNCTION, AND SUBCELLULAR LOCATION.
RX MEDLINE=21564328; PubMed=11707530;
RA Huang H.-K., Forsburg S.L., John U.P., O'Connell M.J., Hunter T.;
RT "Isolation and characterization of the Pin1/Ess1p homologue in
RT Schizosaccharomyces pombe.";
RL J. Cell Sci. 114:3779-3788 (2001).
CC -!- FUNCTION: Has a role in the G1/S stage transition of mitosis
CC where it is involved in the dephosphorylation of cdc25 and weel.
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the ppiC/parvulin rotamase family.
CC -!- SIMILARITY: Contains 1 WW domain.
```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL031535; CAA20742.1; --
DR PIR; T41093; T41093.
DR HSP; Q13526; LPIN.
DR GeneDB_SPombe; SPCC16C4.03; --.
DR InterPro; IPR000297; Rotamase.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00639; Rotamase; 1.
DR Pfam; PF00397; WW; 1.
DR SMART; SM00456; WW; 1.
DR PROSITE; PS01096; PPIC_PPIASE_1; FALSE_NEG.
DR PROSITE; PS01198; PPIC_PPIASE_2; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; FALSE_NEG.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
KW Isomerase; Rotamase; Nuclear protein; Mitosis.
FT DOMAIN 4 38
FT DOMAIN 64 175
FT SEQUENCE 175 AA; 19773 MW; ABA637835471BD25 CRC64;
Query Match 45.6%; Score 388.5; DB 1; Length 175;
Best Local Similarity 48.5%; Pred. No. 4.3e-29;
Matches 82; Conservative 21; Mismatches 51; Indels 15; Gaps 2;
QY 7 LPFGWEKMSRSSGRVYFNHITNASQWERPSGNSGGK-----NGQGEF 52
Db 6 LPKFWIKISRSRPRVFNTTHESLWPPAATDMAALKKFANLQESVTPTEANSP 65
QY 53 ARVCSSHLVVKHSRPPSSRWQKITTKEALELINGYIOKIKSGEEDFESLASPSD 112
Db 66 -KIPASHLLVKKSRPPSSRWKEHITKSKKAEARKLAHYEQLLKSGSVSMHDLAKESD 124
QY 113 CSSNARGDLGAFSGQKQPFEDASFAIRTGEMSGPVFTDSGHIILR 161
Db 125 CSSARRGGELGEGFDEMQKQPFEDAAFAFKGEGISGVVETSSGHIILR 173
RESULT 6
ID ESSL_YEAST STANDARD; PRT; 190 AA.
AC P22636;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ESSL protein (Processing/termination factor 1).
GN ESSL OR PTF1 OR YJR017C OR J1452.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89189095; PubMed=2648698;
RA Hanes S.D., Shank P.R., Bostian K.A.;
RT "Sequence and mutational analysis of ESSL, a gene essential for
RT growth in Saccharomyces cerevisiae.";
RL Yeast 5:55-72(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=DH484;
RC MEDLINE=95300974; PubMed=7781779;
RA Hani J., Stumpf G., Domdey H.;
RT "Ptf1 encodes an essential protein in Saccharomyces cerevisiae, which
RT shows strong homology with a new putative family of ppiases.";
RL FEBS Lett. 365:198-202(1995).
RN [3]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=S288c/F1679;
RA de Haan M., Smits P.H.M., Grivell L.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Essential for growth in Saccharomyces cerevisiae. May be
CC involved in cytokinesis or in cell separation.
CC -!- SIMILARITY: Contains 1 WW domain.
CC -!- SIMILARITY: Belongs to the ppiC/parvulin rotamase family.
CC -!- CAUTION: It is uncertain whether Met-1 or Met-21 is the initiator.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X85972; CAA59961.1; ALT_INIT.
DR EMBL; X87611; CAA60941.1; --.
DR EMBL; Z49517; CAA89541.1; --.
DR PIR; S52764; S52764.
DR HSP; Q13526; LPIN.
DR Germonline; 141854; --.
DR SGD; S0003778; ESSL.
DR GO; GO:0005634; C:nucleus; IPI.
DR GO; GO:0003755; F:peptidyl-prolyl cis-trans isomerase activity; IDA.
DR GO; GO:0006397; P:mRNA processing; IMP.
DR GO; GO:0042326; P:negative regulation of phosphorylation; IDA.
DR InterPro; IPR000297; Rotamase.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00639; Rotamase; 1.
DR Pfam; PF00397; WW; 1.
DR SMART; SM00456; WW; 1.
DR PROSITE; PS01096; PPIC_PPIASE_1; 1.
DR PROSITE; PS01198; PPIC_PPIASE_2; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
KW Isomerase; Rotamase.
FT DOMAIN 29 62
FT DOMAIN 77 190
FT CONFLICT 28 28 R -> S (IN REF. 1).
FT CONFLICT 37 37 V -> A (IN REF. 1).
FT CONFLICT 147 190
FT SEQUENCE 190 AA; 21733 MW; 93E449E3D7B4D989 CRC64;
Query Match 41.8%; Score 356.5; DB 1; Length 190;
Best Local Similarity 46.3%; Pred. No. 4.4e-26;
Matches 75; Conservative 24; Mismatches 61; Indels 3; Gaps 2;
QY 1 MADEKLPPGWEKMSRSSGRVYFNHITNASQWERPSG-NSSSGKNGQGEFARVRCSH 59
Db 25 VASRTGLPTWTVRYSKSKREYFFNPETKHSQWEEPEGTNKQLKHLRHDHVRVRLCH 84
QY 60 LLVKHSQSRPPSSRWQKITTKEALELINGYIOKIKSGEEDFESLASPSDCSSAK 117
Db 85 ILIKHDSRRPASHRSSENTISKQDATDELTKLTITRLDDDDSKTNSPEALAKERSDCSSYK 144
QY 118 ARGDLGAFSGQKQPFEDASFAIRTGEMSGPVFTDSGHIILR 161
Db 145 RGDLGWFGRGEMQSPFEDAAFAQLKQGEVSDIVESGSGVHVIR 188
RESULT 7
PIN1_ARATH STANDARD; PRT; 119 AA.
ID PIN1_ARATH
AC Q9SL42; Q42334;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase 1 (EC 5.2.1.8) (Rotamase Pin1)

```









```

Db 131 BGLKGTIRASHILVADK-----KTAEEV-----EKLLKGEK-FEDLAK 168
Qy 109 QFSDCSAKARAGDLGARSR-GOMOKPFEDASFAIRTGEMSGPVFTDSGIHILATE 163
Db 169 EYSTDSASKSGDGLGWFAKEGQMDFTFKAAPKUKLTGEVSDPVKTQYGYHIKKTE 224

RESULT 12
PIN4 HUMAN
ID PIN4_HUMAN STANDARD; PRT; 131 AA.
AC Q9Y237;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4 (EC 5.2.1.8)
DE (Rotamase Pin4) (PPIase Pin4) (Parvulin 14) (Par14) (Peptidyl-prolyl
DE cis/trans isomerase EPVH) (hPar14).
GN PIN4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=99198957; PubMed=10100858;
RA Uchida T., Fujimori F., Tradler T., Fischer G., Rahfeld J.-U.;
RT "Identification and characterization of a 14 kDa human protein as a
RT novel parvulin-like peptidyl prolyl cis/trans isomerase.";
RL FEBS Lett. 446:278-282(1999).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=92294627; PubMed=10364457;
RA Rulten S.L., Thorpe J.R., Kay J.E.;
RT "Identification of eukaryotic parvulin homologues: a new subfamily of
RT peptidylprolyl cis-trans isomerases.";
RL Biochem. Biophys. Res. Commun. 259:557-562(1999).
[3]
RN SEQUENCE FROM N.A.
RX MEDLINE=922388257; PubMed=12477932;
RA Klausner R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Besak S.A., McEwan P.J., McKernan K.J., Shevchenko Y., Bouffard G.G.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Kryewski M.I., Skalska U., Smalhus D.B.,
RA Schnerch A., Schein J.Y., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[4]
RN STRUCTURE BY NMR OF 36-131.
RX MEDLINE=20425097; PubMed=10966801;
RA Sekerina E., Rahfeld J.-U., Mueller J., Fanghaenel J., Rascher C.,
RA Fischer G., Bayer P.;
RT "NMR solution structure of hPar14 reveals similarity to the peptidyl
RT prolyl cis/trans isomerase domain of the mitotic regulator hPini but
RT indicates a different functionality of the protein.";
RL J. Mol. Biol. 301:1003-1017(2000).
[5]
RN STRUCTURE BY NMR OF 28-131.
RX MEDLINE=21103679; PubMed=11162102;
RA Terada T., Shirouzu M., Fukumori Y., Fujimori F., Ito Y., Kigawa T.,

```

```

RA Yokoyama S., Uchida T.;
RT "Solution structure of the human parvulin-like peptidyl prolyl
RT cis/trans isomerase, hPar14.";
J. Mol. Biol. 305:917-926(2001).
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: Belongs to the pPIC/parvulin rotamase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB009690; BAA82320.1; -
DR EMBL; AF143096; AAD27893.1; -
DR EMBL; BC005234; AAH05234.1; -
DR Genew; HGNC:8992; PIN4.
DR MIM; 300252; -
DR PDB; 1EQ3; 04-APR-01.
DR PDB; 1FJD; 08-AUG-01.
DR GO; GO:0005759; C:mitochondrial matrix; TAS.
DR GO; GO:0006457; P:protein folding; TAS.
DR InterPro; IPR000297; Rotamase.
DR Pfam; PF00639; Rotamase; 1.
DR PROSITE; PS01096; PPI_PPIASE_1; FALSE_NEG.
DR PROSITE; PS01198; PPI_PPIASE_2; 1.
DR Isomerase; Rotamase; 3D-structure.
FT DOMAIN 35 129 PPI_C.
SQ SEQUENCE 131 AA; 13910 MW; 787C15BDB0701258 CRC64;

Query Match 17.8%; Score 152; DB 1; Length 131;
Best Local Similarity 34.0%; Pred. No. 2.6e-07;
Matches 51; Conservative 14; Mismatches 33; Indels 52; Gaps 9;

Qy 37 PSGNSSSG-----GKNGQGEPAVRCSHLIV-KHSQRSPSSWRQE 76
Db 3 PKGSGSGKAGKGAASGSADKKAQPGKGGNAVKR--HILCEKHG----- 49
Qy 77 KITRTKEALELINGYTIKISGEEDPEFSLASQFDCSSAKARGLDGAFRGOMOKPFED 136
Db 50 KIM-----EAME-----KLKSGMR-FNEVAQYSE-DKARQGDGLGWMTGSMVGPQE 96
Qy 137 ASFAIRTGEMSGPVFTDS-----GIHIL 160
Db 97 AAFALPVSGMDKPVFTDPVVKTRFGYHIIM 126

RESULT 13
PIN4 MOUSE
ID PIN4_MOUSE STANDARD; PRT; 131 AA.
AC Q9CWW6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4 (EC 5.2.1.8)
DE (Rotamase Pin4) (PPIase Pin4).
GN PIN4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=C57BY/6J; TISSUE=Embryonic stem cells;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

```



```

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsumoto Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wegner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Bojelli D., Bojunga N., Carninci P., Fujita M., Gariboldi M.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Nombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Hayashizaki Y., Yoshida K., Hasegawa Y., Kawaji H., Kohatsu S.,
RA Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -1- SIMILARITY: Belongs to the pp1C/parvulin rotamase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AK010338; BAB26863.1; -.
CC HSPSP; Q9V237; 1803.
CC MGD; MGI:1916963; Pin4.
CC InterPro; IPR000297; Rotamase.
CC Pfam; PF00639; Rotamase; 1.
CC PROSITE; PS01096; PPIC_PPIASE_1; FALSE_NEG.
CC PROSITE; PS01096; PPIC_PPIASE_2; 1.
CC KW Isomerase; Rotamase.
CC FT DOMAIN 35 129 PPTC.
CC SEQUENCE 131 AA; 13815 MW; 290BDEF72DC69CA6 CRC64;
CC
CC Query Match 17.8%; Score 152; DB 1; Length 131;
CC Best Local Similarity 34.0%; Pred. No. 2.6e-07;
CC Matches 51; Conservative 14; Mismatches 33; Indels 52; Gaps 9;
CC
CC QY 37 PSQNSSSG-----GKNGQGEAPARVRCSHLLV-KHSQSRPSSWRQE 76
CC Db 3 PKGSGSGKGGKGAAGSDSADK3QGPKGGNVAVKVR--HTLCBKHG-----49
CC
CC QY 77 KITRTKEALELINGVIQIKSGEEDFESLASQFDCSSAKARGDILGAFSRGOMQKPFED 136
CC Db 50 KIM-----EAME-----KLKSGMR-FSEVATQISE-DRARQGGDILGWMTRGSMVGFQF 96
CC
CC QY 137 ASFALRTGMSGVPVFTDS-----GTHILL 160
CC Db 97 AAFALPVGMDKPVFTDPVPVKTKFGVHIIM 126
CC
CC RESULT 14
CC CBF2_CAMJ2 STANDARD; PRT; 273 AA.
CC AC Q46105;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Cell binding factor 2 precursor (Major antigen pab4A).
CC GN CBF2 OR PEB4A OR CU0596.
CC OS Campylobacter jejuni.
CC OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
CC OC Campylobacteraceae; Campylobacter.
CC OX NCBI_TaxID=197;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=ATCC 55026 / 81-176;
CC RX MEDLINE=96099687; PubMed=8525063;

```

```

RA Buruoca C., Frenaux C., Pei Z., Tummuru M., Blaser M.J.,
RA Cenatiempo Y., Fauchere J.L.;
RT Nucleotide sequence and characterization of pab4A encoding an
RT antigenic protein in Campylobacter jejuni."
RL Res. Microbiol. 146:467-476(1995).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Baeham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA Jagels K., Kariyeh A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -1- SIMILARITY: Belongs to the pp1C/parvulin rotamase family.
CC STRONG, TO H.FYLORI HP0175.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X84703; CA459175.1; -.
CC DR EMBL; AL139075; CAB75232.1; -.
CC DR PIR; S52412; S52412.
CC DR HSPSP; Q13526; 1PIN.
CC InterPro; IPR000297; Rotamase.
CC Pfam; PF00639; Rotamase; 1.
CC PROSITE; PS01096; PPIC_PPIASE_1; 1.
CC PROSITE; PS01096; PPIC_PPIASE_2; 1.
CC KW Isomerase; Rotamase; Antigen; Signal; Complete proteome.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 273 CELL BINDING FACTOR 2.
CC FT DOMAIN 131 228 PPIC.
CC SEQUENCE 273 AA; 30518 MW; 1C014658B8BC4E39 CRC64;
CC
CC Query Match 17.5%; Score 149; DB 1; Length 273;
CC Best Local Similarity 36.5%; Pred. No. 1.2e-06;
CC Matches 42; Conservative 20; Mismatches 31; Indels 22; Gaps 5;
CC
CC QY 51 EPARVRCSHLLVKSQSRPSSWRQEKITRTKEALELINGVIQIKSGEED--PESLAS 108
CC Db 130 KPARVQAKHILV-----ATEKEAKDIIN-ELKGLKGLKELDAKFSELAK 171
CC
CC QY 109 QFS-DCSSAKARGDILGAFSRGOMQKPFEDASFLRTGEM-SGPFVTDSCIHILR 161
CC Db 172 EKSIDPGSKNGGELGWFEDQSTWVPFTDAAPALKNGTITTPVKTNGYHILK 226
CC
CC RESULT 15
CC PPIC_ECOLI STANDARD; PRT; 92 AA.
CC AC P39159;
CC DT 01-FEB-1995 (Rel. 31, Created)
CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Peptidyl-prolyl cis-trans isomerase C (EC 5.2.1.8) (PPIase C)
CC DE (Rotamase C) (parvulin).
CC GN PPIC OR PARVA OR B3775 OR C4697 OR Z5286 OR ECS4709.
CC OS Escherichia coli,
CC OS Escherichia coli O6, and
CC OS Escherichia coli O157:H7.
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC OC Enterobacteriaceae; Escherichia.
CC OX NCBI_TaxID=562, 217992, 83334;
CC RN [1]
CC RP SEQUENCE FROM N.A., AND SEQUENCE.

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:24 ; Search time 3445.82 Seconds  
(without alignments)  
8614.203 Million cell updates/sec

Title: U49070

Perfect score: 994

Sequence: 1 TGTGCGCCAGCACCTCGAGG.....CCCAATTAAACCCAGACCA 994

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_fod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	895.2	90.1	1035	9	AL526064
C 2	891.4	89.7	988	9	AL526011
C 3	878.8	88.4	950	12	BG744866
4	859	86.4	1076	13	BX436611

5	848.6	85.4	923	12	BI829812
6	817.2	82.2	1201	13	EX463403
C 7	806.2	81.1	1097	9	AL529668
8	805.4	81.0	868	9	AL533491
C 9	799.8	80.5	1039	13	BX436610
10	792.8	79.8	1201	13	EX421727
11	784.2	78.9	864	13	BU557229
12	781	78.6	928	13	BQ678260
13	776.8	78.1	833	12	BG744294
14	775.6	78.0	964	12	BI753931
15	772.2	77.7	874	10	BE797127
16	768.8	77.3	819	12	BI871560
17	762	76.7	837	12	BI597391
18	757.6	76.2	952	13	BQ948365
19	753.6	75.8	1075	13	BX361953
20	750.2	75.5	772	12	BI598834
21	745.4	75.0	921	12	BI596653
C 22	737.4	74.2	797	14	CD364657
23	735	73.9	954	13	BQ068942
24	729.8	73.4	1089	9	AL529669
25	729.2	73.4	854	14	CA454466
26	726	73.0	938	13	BQ430612
27	725.6	73.0	963	13	BU500536
28	724	72.8	919	13	BU182574
29	716.6	72.1	935	14	CA487282
C 30	716	72.0	765	14	CB321634
C 31	716	72.0	772	14	CD365653
C 32	714.8	71.9	817	10	BF570461
33	714	71.8	982	13	BQ956997
34	712.8	71.7	816	12	BG471419
35	700	70.4	712	12	BI489909
C 36	693.4	69.8	767	13	BU683491
37	682.8	68.7	937	9	AU123122
C 38	677.4	68.1	740	14	CD367366
39	676.4	68.0	1200	10	BF314489
40	674.4	67.8	923	10	BF793524
41	671	67.5	728	9	AU123171
42	669.2	67.3	726	10	BE893240
43	669.2	67.3	742	12	BG395172
44	665.8	67.0	680	12	BI856005
45	665.6	67.0	511	14	CA487426

## ALIGNMENTS

RESULT 1  
AL526064  
LOCUS  
DEFINITION  
AL526064 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
cDNA clone CS0DC015YK14 5-PRIME, mRNA sequence.  
1035 bp mRNA linear EST 23-MAY-2003  
ACCESSION  
AL526064  
VERSION  
AL526064.2 GI:31063925  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
REFERENCE  
1 (bases 1 to 1035)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE  
Full-length cDNA libraries and normalization  
JOURNAL  
Unpublished (2001)  
COMMENT  
On Feb 13, 2001 this sequence version replaced gi:12789557.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 9850.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DC015BF07QF1&cluster=9850.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0DC015BF07QPL.

# FEATURES

location/Qualifiers  
1..1035  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DC015YK14"  
/tissue\_type="NEUROBLASTOMA"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 90.1%; Score 895.2; DB 9; Length 1035;  
Best Local Similarity 97.9%; Pred. No. 1.4e-170;  
Matches 946; Conservative 3; Mismatches 11; Indels 6; Gaps 4;

29 CGGACGAGGAGAACTGCGCCCGCTGGGAGAGCGCATGAGCGCGAGCTCAGGCCGAG 88  
30 CGGACGAGGAGAACTGCGCCCGCTGGGAGAGCGCATGAGCGCGAGCTCAGGCCGAG 128  
31 TGTTACTTCTCAACCATCATCTAAGCCAGCCAGCTGGGAGCGGCCAGCGGCAACAGCA 148  
32 TGTTACTTCTCAACCATCATCTAAGCCAGCCAGCTGGGAGCGGCCAGCGGCAACAGCA 188  
33 GCAGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 208  
34 GCAGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 248  
35 TGAAGCAGAGCAGTCAAGCGGCGCCCTGCTGCTGGCGGAGAGGAGGAGGAGGAGGAGG 268  
36 TGAAGCAGAGCAGTCAAGCGGCGCCCTGCTGCTGGCGGAGAGGAGGAGGAGGAGGAGG 308  
37 AGGAGGAGGCGCTGGAGCTGATCAAGCGCTCATCTCAGAGGATCAAGTTCGGGAGGAGG 328  
38 AGGAGGAGGCGCTGGAGCTGATCAAGCGCTCATCTCAGAGGATCAAGTTCGGGAGGAGG 368  
39 ACTTTGAGTCTGCGGCTCAGGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 388  
40 ACTTTGAGTCTGCGGCTCAGGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 428  
41 TGGGTGCTTTCAGCAGAGGTCAGATGCAAGAGCCATTGAAGAGCGCTTCGTTGCGCTGC 448  
42 TGGGTGCTTTCAGCAGAGGTCAGATGCAAGAGCCATTGAAGAGCGCTTCGTTGCGCTGC 488  
43 GGACGGGGAGATGAGCGGGCCCGTGTTCAGGAGTTCGGGATTCACATCATCTCTCCGCA 508  
44 GGACGGGGAGATGAGCGGGCCCGTGTTCAGGAGTTCGGGATTCACATCATCTCTCCGCA 548  
45 CTGAGTCAGGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 568  
46 CTGAGTCAGGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 608  
47 CAGTCTCCCTTCGCGCGCAGCAGTGGCGGAAACCCCGCCACTCTCCCTGCGCAGCGTCACACA 628  
48 CAGTCTCCCTTCGCGCGCAGCAGTGGCGGAAACCCCGCCACTCTCCCTGCGCAGCGTCACACA 668  
49 GTATTATTGTTCCCAATGGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 688  
50 GTATTATTGTTCCCAATGGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 726  
51 CCCTCTCCCTTCCTCCAGTTGGGCTGCGAGCGGCGAGATTCTCCCTTAAGGAGATTG 748  
52 CCCTCTCCCTTCCTCCAGTTGGGCTGCGAGCGGCGAGATTCTCCCTTAAGGAGATTG 785  
53 ACTTCAGCAGGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 808  
54 ACTTCAGCAGGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 845  
55 AGAAGAGGCTGTGTACAGAGAGCGCCCGCGCTGTCCCCCGCCAGGAGGAGGAGGAGGAG 868

846 AGAAGAGGCTGTGTACAGAGAGCGCCCGCTGTCCCCCGAGGTCTGCGAGGAGGAGGAG 905  
847 AGAAGAGGCTGTGTACAGAGAGCGCCCGCTGTCCCCCGAGGTCTGCGAGGAGGAGGAG 928  
848 GGGCGGAAATGTTCTAGTTAGGCGCAAGCTCTCTGTTGTTGAGTTCGCAAGAGGAGGAG 965  
849 GGGCGGAAATGTTCTAGTTAGGCGCAAGCTCTCTGTTGTTGAGTTCGCAAGAGGAGGAG 986  
850 TGGCG--CAGGCATGGGCGCTCTGAGCAACTGTGAGACCCCTTTCACCCCGCAATTAAC 1024  
851 TGGCGCCAGCAGTGGGCGCTCTGAGC-ACTGTGASAMCCTTTCACCCCGCAATTAAC 1024  
852 CAGAAC 992  
853 GAACAC 1030

RESULT 2  
AL526011/c  
LOCUS  
DEFINITION AL526011 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
CDNA clone CS0DC015YK14 3-PRIME, mRNA sequence.  
ACCESSION AL526011  
VERSION AL526011.2 GI:31063874  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 988)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On Feb 13, 2001 this sequence version replaced gi:12789504.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 9850.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DC015BF07NP1&cluster=9850.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DC015BF07NP1.

## FEATURES

source  
1..988  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DC015YK14"  
/tissue\_type="NEUROBLASTOMA"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 89.7%; Score 891.4; DB 9; Length 988;  
Best Local Similarity 98.2%; Pred. No. 8.2e-170;  
Matches 907; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

29 CGGACGAGGAGAACTGCGCCCGCTGGGAGAGCGCATGAGCGCGAGCTCAGGCCGAG 88  
30 CGGACGAGGAGAACTGCGCCCGCTGGGAGAGCGCATGAGCGCGAGCTCAGGCCGAG 901  
31 TGTTACTTCTCAACCATCATCTAAGCCAGCCAGCTGGGAGCTGGCGCCAGCGGCAACAGCA 148  
32 TGTTACTTCTCAACCATCATCTAAGCCAGCCAGCTGGGAGCTGGCGCCAGCGGCAACAGCA 841  
33 GCAGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 208







```

ACCESSION      BI829812
VERSION        BI829812.1  GI:15941362
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens

REFERENCE      1 (bases 1 to 923)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgabs@email.nih.gov
              Tissue procurement: Life Technologies, Inc.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM1426 row: n column: 21
              High quality sequence stop: 861.
              Location/Qualifiers
                1..923
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:5171540"
                  /tissue_type="medulla"
                  /lab_host="DH10B"
                  /clone_lib="NIH MGC 119"
                  /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
                  site 2: EcoRV (destroyed); RNA source normal medulla from
                  anonymous male age 27. Library is oligo-dT primed and
                  directionally cloned (EcoRV site is destroyed upon
                  cloning). Average insert size 1.3 kb, insert size range
                  0.9-3 kb. Library is normalized and enriched for
                  full-length clones and was constructed by C. Gruber
                  (Invitrogen). Research Genetics tracking code 013. Note:
                  this is a NIH_MGC Library."

ORIGIN
Query Match      85.4%; Score 848.6; DB 12; Length 923;
Best Local Similarity 98.4%; Pred. No. 3.6e-161;
Matches 910; Conservative 0; Mismatches 19; Indels 6; Gaps 5;

QY      23  AGATCGCGACGAGGAGAGCTGCCCGCGCTGGGAGAGCGCATGAGCCGAGCTCAG 82
Db      |||||||
QY      83  GCGGAGTGTAATCTTCAACCATCATCACTAAGCCAGCCAGTGCGGAGCGGCCAGCGGCA 142
Db      |||||||
QY      61  GCGGAGTGTAATCTTCAACCATCATCACTAAGCCAGCCAGTGCGGAGCGGCCAGCGGCA 120
Db      |||||||
QY      143  ACAGCAGCAGTGCTGTCACAAAACGCGCAGCGGAGCGCTGCGAGGCTGCGTGCACAC 202
Db      |||||||
QY      121  ACAGCAGCAGTGCTGTCACAAAACGCGCAGCGGAGCGCTGCGAGGCTGCGTGCACAC 180
Db      |||||||
QY      203  TCCTGCTGAACACACGCGCAGTCACGCGCGCGCTGCTGCTGCGCGCAGGAGATCACCC 262
Db      |||||||
QY      181  TCTGCTGGAAGCAGACCGCAGTCACGCGCGCGCTGCTGCTGCGCGCAGGAGATCACCC 240
Db      |||||||
QY      263  GGACCAAGGAGGAGCGCCCTGGAGCTGATCAAGCGCTACATCCAGAGATCAAGTCGGGAG 322
Db      |||||||
QY      241  GGACCAAGGAGGAGCGCCCTGGAGCTGATCAAGCGCTACATCCAGAGATCAAGTCGGGAG 300
Db      |||||||
QY      323  AGGAGCACTTTGAGCTCTGCGGCTCAAGTTAGCGACTGCGAGCTCAGCCAGCGCCAGGG 382
Db      |||||||
QY      301  AGGAGCACTTTGAGCTCTGCGGCTCAAGTTAGCGACTGCGAGCTCAGCCAGCGCCAGGG 360
Db      |||||||
QY      383  GAGAGCTCGGCTTTCAGCAGAGGTCAGATCAGAGGACCATTTGAGAGCGCTCGCTTG 442
Db      |||||||
QY      361  GAGAGCTCGGCTTTCAGCAGAGGTCAGATCAGAGGACCATTTGAGAGCGCTCGCTTG 420

```

```

QY      443  CGCTGCGGACGCGGCGAGATGAGCGGGCCCGTGTTCAGGATTCGGGATCCACATCATCC 502
Db      |||||||
QY      421  CGCTGCGGACGCGGCGAGATGAGCGGGCCCGTGTTCAGGATTCGGGATCCACATCATCC 480
Db      |||||||
QY      503  TCCGCACTGAGTGAGGCTGGGAGCGCCAGCCCTGGCCCTCGGGGAGGCGGCGGCTAG 562
Db      |||||||
QY      481  TCCGCACTGAGTGAGGCTGGGAGCGCCAGCCCTGGCCCTCGGGGAGGCGGCGGCTAG 540
Db      |||||||
QY      563  GCGGCGCAGCTCCCTTGCCTCCGCGCAGCAGTGCAGAACCCCGCCACTCCCTGCCACCGT 622
Db      |||||||
QY      541  GCGGCGCAGCTCCCTTGCCTCCGCGCAGCAGTGCAGAACCCCGCCACTCCCTGCCACCGT 600
Db      |||||||
QY      623  CACACAGTATTTATTTGTTCCACATGGCTGGGAGGGGGCCCTTCCAGATTTGGGGGCCCT 682
Db      |||||||
QY      601  CACACAGTATTTATTTGTTCCACATGGCTGGGAGGGGGCCCTTCCAGATTTGGGGGCCCT 660
Db      |||||||
QY      683  GGGGTCCCACTCCCTGCTCCATCCAGTTGGGGCTGCGACCGCCAGATTTCCCTTAAG 742
Db      |||||||
QY      661  GGGGTCCCACTCCCTGCTCCATCCCAAGTTGGGGCTGCGACCGCCAGATTTCCCTTAAG 720
Db      |||||||
QY      743  GAATGACTTTCAGCAGGGGTGGAGGCTCCAGACCCAGGCGAGTGTGGGAGGGG-- 800
Db      |||||||
QY      721  GAATGACTTTCAGCAGGGGTGGAGGCT--CCAGACCCAGGCGAGTGTGGGAGGGGGT 779
Db      |||||||
QY      801  TGTTCCAAAGAGAA--GSCCTGTGTACAGAGCGCCCGCTGTCCCGCCAGGTGTGGAGG 859
Db      |||||||
QY      780  TGTTCCAAAGAGAAAGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 838
Db      |||||||
QY      860  CAGACTCGAGGGCGGAAATGTTTCTAGTTAGGCGCAGC--TCCTCTGTTCAGTCGCAAGG 918
Db      |||||||
QY      839  CAGACTCGAGGGCGGAAATGTTTCTAGTTAGGCGCAGCTTCTCTGTTCAGTCGCAAGG 898
Db      |||||||
QY      919  TGAACACTCATGCGGAGCCATCGG 943
Db      |||||||
QY      899  TGAACACTCATGCGGAGCCATCGG 923
Db      |||||||

RESULT 6
LOCUS      BX463403 1201 bp mRNA linear EST 22-MAY-2003
DEFINITION BX463403 Homo sapiens ADULT BRAIN Homo sapiens cDNA clone
            CS0DN001YN02 5-PRIME, mRNA sequence.
ACCESSION  BX463403
VERSION    BX463403.1
KEYWORDS   GI:31021526
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1201)
AUTHORS    Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished (2001)
COMMENT    Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 9850.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DN001DG01QP1&cluster=9850.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Paraday Avenue Genoscope sequence ID : CS0DN001DG01QP1.
            Location/Qualifiers
              1..1201
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CS0DN001YN02"
                /tissue_type="ADULT BRAIN"
FEATURES
source

```

```

/dev stage="adult"
/clone.lib="Homo sapiens ADULT BRAIN"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN
Query Match      82.2%; Score 817.2; DB 13; Length 1201;
Best Local Similarity 99.3%; Pred. No. 9e-155;
Matches 842; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

QY 150 CAGTGTGGTGGCAAAACGGCGAGGGGAGCTGCGAGGGTCCGCTGCTGCACCTGCTGCT 209
DB 60 CCGGATGGCAAAACGGCGAGGGGAGCTGCGAGGGTCCGCTGCTGCACCTGCTGCT 119

QY 210 GAAGCACACCCAGTCAAGCGCGCCCTGCTGCTGGCGGCGAGGAAGATCACCGGACCAA 269
DB 120 GAAGCACACCCAGTCAAGCGCGCCCTGCTGCTGGCGGCGAGGAAGATCACCGGACCAA 179

QY 270 GGAGAGGCCCTGGAGCTGATCAAGCGCTACATCCAGAGATCAAGTGGGAGAGGGA 329
DB 180 GGAGAGGCCCTGGAGCTGATCAAGCGCTACATCCAGAGATCAAGTGGGAGAGGGA 239

QY 330 CTTTGAGTCTCTGGGCTCAAGTTCAGGCTGAGCTGAGCTCAGGCCAGGCGAGGAGACCT 389
DB 240 CTTTGAGTCTCTGGGCTCAAGTTCAGGCTGAGCTGAGCTCAGGCCAGGCGAGGAGACCT 299

QY 390 GGGTGCCTTACAGAGAGTCAAGTCAAGAGCAATTGAGAGCGCTGCTGCTGGCTGGC 449
DB 300 GGGTGCCTTACAGAGAGTCAAGTCAAGAGCAATTGAGAGCGCTGCTGCTGGCTGGC 359

QY 450 GACGGGGAGATGACGGGCGCGTGTTCAGGATTCGGCATCCGATCATCTCTCGGCAC 509
DB 360 GACGGGGAGATGACGGGCGCGTGTTCAGGATTCGGCATCCGATCATCTCTCGGCAC 419

QY 510 TCAGTTCAGGCTGGGAGCCAGCGCTGGCTCGGGGAGGCGAGGCGCGCTAGGCGCGCC 569
DB 420 TGGTGTGGGGTGGGAGCCAGGCTGGCTCGGGGAGGCGAGGCGCGCTAGGCGCGCC 479

QY 570 AGCTCCCTTGCCTGCGGCGAGCGAGTGGCGAAGCCCGGCTGCTGCTGCGGCGAGTCA 629
DB 480 AGCTCCCTTGCCTGCGGCGAGCGAGTGGCGAAGCCCGGCTGCTGCTGCGGCGAGTCA 539

QY 630 TATTTATTTCTCCCAATGGCTGGGAGGGGCGCTTCCAGATTGGGGGCGCTGGGGTCC 689
DB 540 TATTTATTTCTCCCAATGGCTGGGAGGGGCGCTTCCAGATTGGGGGCGCTGGGGTCC 599

QY 690 CCACTCCCTGCTCCATCCCGAGTGGGGCTGCGGACCGCGAGTCTCCCTTAAGGAATTGA 749
DB 600 CCACTCCCTGCTCCATCCCGAGTGGGGCTGCGGACCGCGAGTCTCCCTTAAGGAATTGA 659

QY 750 CTTTACAGAGGGTGGAGGCTCCAGACCCAGGCGAGTGTGGTGGAGGGGTGTTCCAAA 809
DB 660 CTTTACAGAGGGTGGAGGCTCCAGACCCAGGCGAGTGTGGTGGAGGGGTGTTCCAAA 719

QY 810 GAGAGGCGCTGTTGACGAGAGCGCGCGCTGCTCCCGAGTGTCTGGAGGCGAGCTCGAG 869
DB 720 GAGAGGCGCTGTTGACGAGAGCGCGCGCTGCTCCCGAGTGTCTGGAGGCGAGCTCGAG 779

QY 870 GGCAGGAATTGTTTCTAGTTAGGCGCAGCTCCTCTGTTTCACTGCGAAGAGTGAACATCAT 929
DB 780 GGCAGGAATTGTTTCTAGTTAGGCGCAGCTCCTCTGTTTCACTGCGAAGAGTGAACATCAT 839

QY 930 GCGG--CAGGCCATGGGCGCTCTGAGCACTGTGAG--ACCTTTTACCCCCCAATTAAACC 986
DB 840 GCGGCGCAGCATGGGCGCTCTGAGCACTGTGAG--ACCTTTTACCCCCCAATTAAACC 899

QY 987 CAGAACCA 994
DB 900 CAGAACCA 907

```

## RESULT 7

AL529668/c  
LOCUS  
DEFINITION  
AL529668 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens  
CDNA clone CS0DD005YA23 3-PRIME, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1097)

Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On Feb 13, 2001 this sequence version replaced gi:12793161.

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 BVR cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster

9850.f For more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DD005AA12NP1&cluster=9850.f. Contact :

Peng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID : CS0DD005AA12NP1.

Location/Qualifiers

1..1097

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DD005YA23"

/tissue type="NEUROBLASTOMA COT 50-NORMALIZED"

/clone lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 81.1%; Score 806.2; DB 9; Length 1097;  
Best Local Similarity 91.4%; Pred. No. 1.5e-152;  
Matches 884; Conservative 15; Mismatches 39; Indels 29; Gaps 4;

QY 11 CACCTCGAGGAAGATGCGGACGAGGAGAGTGTGGCGGCTGCGGCGCTGGGAGGCGATGA 70  
DB 987 CCGGSRGTGGGAGGNTGGCGGACGAGAGAGTGTGGCGGCTGCTGTTGAAATGCGCTGA 928  
QY 71 GCGCGAGCTCAGCGCGAGTGTACTTCAACACATCACTAAACGCGCAGTGGGAGC 130  
DB 927 GCGCGAGCTCAGCGCGAGTGTACTTCAACACATCACTAAACGCGCAGTGGGAGC 869  
QY 131 GCGCGAGCTCAGCGCGAGTGTGTGGCAAAACGCGGAGGCGGAGCTGCCAGGCTCC 190  
DB 868 GCGCGAGCTCAGCGCGAGTGTGTGGCAAAACGCGGAGGCGGAGCTGCCAGGCTCC 810  
QY 191 GCTGCTGCGACCTGCTGCTGAAGCAGCAGCCAGTTCAGGCGGCGCTGCTGCGGCGAG 250  
DB 809 GCTGCTGCGACCTGCTGCTGAAGCAGCAGCCAGTTCAGGCGGCGCTGCTGCGGCGAG 750  
QY 251 AGAAGATTCACCGGACCAAGGAGGCGCTGCGAGTGTATCAACGCGTACATCCAGAGA 310  
DB 749 AGAAGATTCACCGGACCAAGGAGGCGCTGCGAGTGTATCAACGCGTACATCCAGAGA 690  
QY 311 TCAAGTGGGAGGAGGAGCTTTGAGTCTTGGGCTTCAAGTTCAGGAGTTCAGGCTCAG 370  
DB 689 TCAAGTGGGAGGAGGAGCTTTGAGTCTTGGGCTTCAAGTTCAGGAGTTCAGGAGTTCAG 630  
QY 371 CCAAGGCGGAGGAGGAGCTGCTGCTTTCAGCAG-----A 405



```
Db 629 CCAAGCCAGGAGAGACCTGGGTGCTTCAGCAGAGGTGCGAAGGATGGCCCTCACCA 570
QY 406 GGTTCAGATGAGAGAGCCATTGAGAGCGCTTTCGCTGCGTGGAGAGGGGAGATGAGC 465
Db 569 GGTTCAGATGAGAGAGCCATTGAGAGCCCTTGGTTTCGCTGCGGAGCGGGGAGNTGGG 510
QY 466 GGGCCCGTGTTCACGAGTTCGAGATCCACATCATCTCCGCACTGAGTGAGGAGTGGGGA 525
Db 509 GGGCCCGGKGTTCAGGATTCGAGCATCCACATCATCTCCGCACTGAGTGAGGAGTGGGGA 450
QY 526 GCCCAGGCTGCTGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 585
Db 449 GCCCAGGCTGCTGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 390
QY 586 CCAGCAGAGTGGGCGAGCCCGCCCTCCCTCCGACCGTCCACAGATATTTATTTCCAC 645
Db 389 CCAGCAGAGTGGGCGAGCCCGCCCTCCCTCCGACCGTCCACAGATATTTATTTCCAC 330
QY 646 AATGGCTGGAGAGGGGCGCTTCCAGATTTGGGAGGCGCTTCCGAGTCCCTGTTCCATC 705
Db 329 AATGGCTGGAGAGGGGCGCTTCCAGATTTGGGAGGCGCTTCCGAGTCCCTGTTCCATC 270
QY 706 CCGAGTGGGCGTCCGAGCGGAGATTCCTCCCTTACGAGATTCAGTTCAGCAGGAGTGG 765
Db 269 CCGAGTGGGCGTCCGAGCGGAGATTCCTCCCTTACGAGATTCAGTTCAGCAGGAGTGG 210
QY 766 AGGCTCCAGAGCCAGGCGAGTGGTGGAGGCGGTTCCTCAAGAGAGAGGCGCTGGTTCAG 825
Db 209 AGGCTCCAGAGCCAGGCGAGTGGTGGAGGCGGTTCCTCAAGAGAGAGGCGCTGGTTCAG 150
QY 826 CAGAGCGGCGGCTGCTGCTCCGAGGAGTGGTGGAGGAGTGGTGGAGGAGTGGTGGAG 885
Db 149 CAGAGCGGCGGCTGCTGCTCCGAGGAGTGGTGGAGGAGTGGTGGAGGAGTGGTGGAG 90
QY 886 GTTAGGCGAGGCTGCTGCTGCTGAGTGGAGGAGTGGTGGAGGAGTGGTGGAGGAGTGG 943
Db 89 GTTAGGCGAGGCTGCTGCTGCTGAGTGGAGGAGTGGTGGAGGAGTGGTGGAGGAGTGG 30
QY 944 CCTCTG 950
Db 29 CCTCTG 23
```

```
RESULT 8
AL533491
LOCUS
DEFINITION
AL533491 Homo sapiens ADULT BRAIN Homo sapiens cDNA clone
CDON004YN04 5-PRIME, mRNA sequence.
ACCESSION
AL533491
VERSION
AL533491.2 GI:31260572
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 13, 2001 this sequence version replaced gi:12796984.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9850.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS00N004DG02QP1&cluster=9850.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1500
Faraday Avenue Genoscope sequence ID : CS00N004DG02QP1.
```

```
FEATURES
source
1..868
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS00N004YN04"
/tissue_type="ADULT BRAIN"
/dev stage="adult"
/clone lib="Homo sapiens ADULT BRAIN"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
ORIGIN
Query Match 81.0%; Score 805.4; DB 9; Length 868;
Best Local Similarity 99.1%; Pred. No. 1.9e-152;
Matches 806; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 13 CCTCGAGGAGAGTGGCGAGGAGAGAGTGGCGGCGCTGGGAGAGCGATGAGC 72
Db 56 CCGGATGGAGATGGCGGAGCGAGGAGAGTGGCGGCGCTGGGAGAGCGATGAGC 115
QY 73 CGCAGCTTCAGGCGGAGTGTACTTCAACCAATCACTAACGCCAGCAGTGGAGCGG 132
Db 116 CGCAGCTTCAGGCGGAGTGTACTTCAACCAATCACTAACGCCAGCAGTGGAGCGG 175
QY 133 CCAGCGGCAACAGCAGCAGTGTGGCAAAACCGGCGAGGAGCTGCCAGGCTCCGC 192
Db 176 CCAGCGGCAACAGCAGCAGTGTGGCAAAACCGGCGAGGAGCTGCCAGGCTCCGC 235
QY 193 TGTCTCGCAGCTGTGTGAGCAGCAGCAGTGTGGCAAAACCGGCGAGGAGCTGCCAGGCTCCGC 252
Db 236 TGTCTCGCAGCTGTGTGAGCAGCAGCAGTGTGGCAAAACCGGCGAGGAGCTGCCAGGCTCCGC 295
QY 253 AAGATCACCCGAGCAACAGGAGGAGGCGCTGAGTGTATCAACGCTATCATTCAGAGATC 312
Db 296 AAGATCACCCGAGCAACAGGAGGAGGCGCTGAGTGTATCAACGCTATCATTCAGAGATC 355
QY 313 AAGTCGGAGAGAGGAGCTTTGAGTCTTGGCTTCAAGCTTCAGCAGCTTCAGCTCAGCC 372
Db 356 AAGTCGGAGAGAGGAGCTTTGAGTCTTGGCTTCAAGCTTCAGCAGCTTCAGCTCAGCC 415
QY 373 AAGGCCAGGAGGAGCAGCTGGTGGCTTCAGCAGAGGTCAGATGCGAGGAGGAGGAGGAG 432
Db 416 AAGGCCAGGAGGAGCAGCTGGTGGCTTCAGCAGAGGTCAGATGCGAGGAGGAGGAGGAG 475
QY 433 GCTCGTGTTCGCTGCGGAGCGGAGGAGTGGCGGCGCTGCTTCAAGCTTCAGCTCAGCTC 492
Db 476 GCTCGTGTTCGCTGCGGAGCGGAGGAGTGGCGGCGCTGCTTCAAGCTTCAGCTCAGCTC 535
QY 493 CACATCATCTTCGCTCAGTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 552
Db 536 CACATCATCTTCGCTCAGTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 595
QY 553 GGGCGCTAGGCGGCGGAGCTCCCTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 612
Db 596 GGGCGCTAGGCGGCGGAGCTCCCTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 655
QY 613 CTGCCACCGTTCACAGATTTATTTGTTCCCAATGCTGGAGGAGGAGGAGGAGGAGGAGG 672
Db 656 CTGCCACCGTTCACAGATTTATTTGTTCCCAATGCTGGAGGAGGAGGAGGAGGAGGAGG 715
QY 673 TGGGGGCGCTTCGAGGCTCCCACTCCCTGTTCATCCCGAGTGGGCTGCGAGCGGAGATG 732
Db 716 TGGGGGCGCTTCGAGGCTCCCACTCCCTGTTCATCCCGAGTGGGCTGCGAGCGGAGATG 775
QY 733 CTCCCTTAAGGAATTCAGCTTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 792
Db 776 CTCCCTTAAGGAATTCAGCTTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 835
QY 793 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 825
```





```

Db      305 AGATCAAGTGGGAGAGAGAGACTTTGAGTCTCTGGCCCTCAAGTTGACGACTGCAGCT 364
QY      368 CAGCAAGGGCCAGGGAGAGACTTGGTGCCTTTAGCAGAGAGTTCAGATGCAAGGCAATTTG 427
Db      365 CAGCAAGGGCCAGGGAGAGACTTGGTGCCTTTAGCAGAGAGTTCAGATGCAAGGCAATTTG 424
QY      428 AAGAGCGCTCGTTTGGCTGGAGCGGGAGAGATGAGCGGCGCGTGTTCAGGATTCGG 487
Db      425 AAGAGCGCTCGTTTGGCTGGAGCGGGAGAGATGAGCGGCGCGTGTTCAGGATTCGG 484
QY      488 GCATCCACATCATCTCCGCACTGAGTCAAGTGGGAGCGCCAGGCGCTGGGCGCA 547
Db      485 GCATCCACATCATCTCCGCACTGAGTGGGAGCGCCAGGCGCTGGGCGCA 544
QY      548 GGGCAGGGCGCTAGGGCGGCGAGCTCCCTTGGCCCGCAGCGAGTGGCGGAGCCCGCC 607
Db      545 GGGCAGGGCGGCTAGGGCGGCGAGCTCCCTTGGCCCGCAGCGAGTGGCGGAGCCCGCC 604
QY      608 ACTCCCTGCCACCGTCAACAGTATTATTGTTTCCCAATGGCTGGGAGGGGCGCTTC 667
Db      605 ACTCCCTGCCACCGTCAACAGTATTATTGTTTCCCAATGGCTGGGAGGGGCGCTTC 664
QY      668 CAGATTGGGGCGGCTCGGGTCCCACTCCCTGTCCATCCCAATGGGCTGGGAGCGCGCC 727
Db      665 CAGATTGGGGCGGCTCGGGTCCCACTCCCTGTCCATCCCAATGGGCTGGGAGCGCGCC 724
QY      728 AGATTCTCCCTTAAGGAATTCAGTTCAGCAGGGGTGGGAGGCTCCGAGCCAGCGGAGT 787
Db      725 AGATTCTCCCTTAAGGAATTCAGTTCAGCAGGGGTGGGAGGCTCCGAGCCAGCGGAGT 784
QY      788 GTGGT--GGGAGGGGTGTTCCAAAGAGAA--GGGCTGGTCAAGAGAGCGCGCGTGTCC 843
Db      785 GTGGTGGGAGGGGTGTTCCAAAGAGAAAGGCGCTGNTCAGCAGAGAGCGCGCGTGT 844
QY      844 CCCC 847
Db      845 CCCC 848

```

## RESULT 12

BQ678260  
LOCUS AGENCOURT\_841988 NIH\_MGC\_112 Homo sapiens cDNA clone IMAGE:6274055  
5', mRNA sequence.

ACCESSION BQ678260

VERSION BQ678260.1 GI:21790939

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Bukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 928)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabs@mail.nih.gov](mailto:cgabs@mail.nih.gov)

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.lnl.gov>

Plate: L10M2454 row: p column: 24

High quality sequence stop: 685.

## FEATURES

source

```

/tissue type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"
/notes="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCCACGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

```

## ORIGIN

```

Query Match      78.6%; Score 781; DB 13; Length 928;
Best Local Similarity 96.5%; Pred. No. 1.7e-147;
Matches 831; Conservative 0; Mismatches 25; Indels 5; Gaps 3;
QY 28 GCGCAGCAGAGAGAGTGCCTCCCGGCTGGGAGAGGCGCATGAGCCGCGAGCTCAGGCGGA 87
Db 1 GCGCAGCAGAGAGAGTGCCTCCCGGCTGGGAGAGGCGCATGAGCCGCGAGCTCAGGCGGA 60
QY 88 GTGTACTACTTCAACCAATCATCTAAAGCCAGAGTGGGAGCGGCCAGCGGCAACAGC 147
Db 61 GTGTACTACTTCAACCAATCATCTAAAGCCAGAGTGGGAGCGGCCAGCGGCAACAGC 120
QY 148 AGCAGTGTGTCGCAAAAGCGGCGAGGCGGAGCTCCAGGGTCCGCTGCTCCACCTGCTG 207
Db 121 AGCAGTGTGTCGCAAAAGCGGCGAGGCGGAGCTCCAGGGTCCGCTGCTCCACCTGCTG 180
QY 208 GTGAAGCAGCAGCCAGTCACTCAGCGCGCCCTCGCTCTGGGCGCAGAGAGAGATCACCGGAGCC 267
Db 181 GTGAAGCAGCAGCCAGTCACTCAGCGCGCCCTCGCTCTGGGCGCAGAGAGATCACCGGAGCC 240
QY 268 AAGCAGAGGCGGCTGGAGCTGATCAAGCGGTACATCGAGAGAGTCAAGTCCGGAGAGAG 327
Db 241 AAGCAGAGGCGGCTGGAGCTGATCAAGCGGTACATCGAGAGAGTCAAGTCCGGAGAGAG 300
QY 328 GACTTTCAGTCTCTGGGCTCAGAGTTCAGGCGACTGAGCTCAGCCAGAGGCGCAGGGAGAG 387
Db 301 GACTTTCAGTCTCTGGGCTCAGAGTTCAGGCGACTGAGCTCAGCCAGAGGCGCAGGGAGAG 360
QY 388 CTGGGTCTCTTACAGAGAGTTCAGATCAGAGAGCCATTTGAAGAGCGCTCGTTTGGGCTG 447
Db 361 CTGGGTCTCTTACAGAGAGTTCAGATCAGAGAGCCATTTGAAGAGCGCTCGTTTGGGCTG 420
QY 448 CGGACGGGGGAGAGTGGCGGGCGGCTTCAAGAGTTCGGGATCCCAATCATCTCTCCGCG 507
Db 421 CGGACGGGGGAGAGTGGCGGGCGGCTTCAAGAGTTCGGGATCCCAATCATCTCTCCGCG 480
QY 508 ACTGAGTGGGGTGGGAGAGCCAGGCGCTGGGCGCAGGGCGAGGCGGCTAGGCGGG 567
Db 481 ACTGAGTGGGGTGGGAGAGCCAGGCGCTGGGCGCAGGGCGAGGCGGCTAGGCGGG 540
QY 568 CAGCTCCCGCTTGCCTCCAGAGTGGCGGAGAAACCCCGCAGCTCCCTGCGACCGCTGACAC 627
Db 541 CAGCTCCCGCTTGCCTCCAGAGTGGCGGAGAAACCCCGCAGCTCCCTGCGACCGCTGACAC 600
QY 628 AGTATTTATTGTTCCCAATGCTGGAGGGGCGCTCCAGATTCGGGCGGCTGGGCTG 687
Db 601 AGTATTTATTGTTCCCAATGCTGGAGGGGCGCTCCAGATTCGGGCGGCTGGGCTG 660
QY 688 CCCCACCTCCCTGTCATCCCGAGTTGGGCTGGGAGCCCGCAGAGTTCTCCCTTAAGGAATT 747
Db 661 CCCCACCTCCCTGTCATCCCGAGTTGGGCTGGGAGCCCGCAGAGTTCTCCCTTAAGGAATT 720
QY 748 GACTTCAGAGGGGTGGGAGGCTCCCGAGAGCCAGGGGAGTGTGGT--GGAGGGGTGTTC 805
Db 721 GACTTCAGAGGGGTGGGAGGCTCCCGAGAGCCAGGGGAGTGTGGT--GGAGGGGTGTTC 780
QY 806 CAAAGAGAGAGGCGCTGGT--CAGCAGAGCGCCCGGCTGTCGCCCGAGGTGTGGAGGCA-G 862
Db 781 CAAAGAGAGAGGCGCTGGT--CAGCAGAGCGCCCGGCTGTCGCCCGGCGGTGGAGGCA 840

```





```
/clone lib="NIH_MGC_114"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: EcoRV (destroyed); RNA source anonymous pool of 6  
male brains, age range 23-27 yo. Library is oligo-dT  
primed and directionally cloned (EcoRV site is destroyed  
upon cloning). Average insert size 1.5 kb, insert size  
range 1-3 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 019. Note:  
this is a NIH_MGC Library."
```

## ORIGIN

```
Query Match 78.0%; Score 775.6; DB 12; Length 964;  
Best Local Similarity 93.5%; Pred. No. 2.1e-146;  
Matches 899; Conservative 0; Mismatches 49; Indels 14; Gaps 8;  
  
QY 26 TGGCGGACGAGGAGAGCTGCGCGCGCTGGGAGAGCGCATGAGCCGAGCTCAGGCC 85  
Db 1 TGGCGGACGAGGAGAGCTGCGCGCGCTGGGAGAGCGCATGAGCCGAGCTCAGGCC 60  
  
QY 86 GAGTGTACTTCAACCACTACATACGCCAGCTGGAGCGGCCCGCCAGCGCAACA 145  
Db 61 GAGTGTACTTCAACCACTACATACGCCAGCTGGAGCGGCCCGCCAGCGCAACA 120  
  
QY 146 GCAGCAGTGGTGCAGGAGAGCGGAGCGCTGCCAGGCTCCGCTGCTGCACCTGC 205  
Db 121 GCAGCAGTGGTGCAGGAGAGCGGAGCGCTGCCAGGCTCCGCTGCTGCACCTGC 180  
  
QY 206 TGGTGAAGCAGACGAGTCAAGCGCGCTCTCTGCGGCGAGGAGATCAACCGGA 265  
Db 181 TGGTGAAGCAGACGAGTCAAGCGCGCTCTCTGCGGCGAGGAGATCAACCGGA 240  
  
QY 266 CCAAGAGAGGAGCTTGGAGCTGATCAACGGCTACATCCAGAGATCAAGTCGGAGAGG 325  
Db 241 CCAAGAGAGGAGCTTGGAGCTGATCAACGGCTACATCCAGAGATCAAGTCGGAGAGG 300  
  
QY 326 AGGCTTTTCACTCTGCGCTCAGAGTTCAGGCTGAGCTCAGCAGAGCGGAGGAG 385  
Db 301 AGGCTTTTCACTCTGCGCTCAGAGTTCAGGCTGAGCTCAGCAGAGCGGAGGAG 360  
  
QY 386 ACCTGGGTGCTTCAAGCAGAGTCAAGTCAAGCAAGCATTTCAGAGCGCTGTTTGGC 445  
Db 361 ACCTGGGTGCTTCAAGCAGAGTCAAGTCAAGCAAGCATTTCAGAGCGCTGTTTGGC 420  
  
QY 446 TGGGACGCGGAGAGTGAAGCGGCGCGTGTTCAGGATTCGGGATCCACATCATCTCC 505  
Db 421 TGGGACGCGGAGAGTGAAGCGGCGCGTGTTCAGGATTCGGGATCCACATCATCTCC 480  
  
QY 506 GCATGTAGTGGGTGGGAGCGCCAGGCGCTGCGCTCGGGGAGGCGAGGCGGCGGCG 564  
Db 481 GCATGTAGTGGGTGGGAGCGCCAGGCGCTGCGCTCGGGGAGGCGAGGCGGCGGCG 540  
  
QY 565 CGGCGAGTCCCTTGCCTGCGGAGCGGAGTGGCGGAGACCCCGGCTGCTGCGACCGTCA 624  
Db 541 CGGCGAGTCCCTTGCCTGCGGAGCGGAGTGGCGGAGACCCCGGCTGCTGCGACCGTCA 600  
  
QY 625 CACAGTATTATTTTCCCAATGGCTGGAGGCGGCGCTTCCAGATTGGGCGGCTGG 684  
Db 601 CACAGTATTATTTTCCCAATGGCTGGAGGCGGCGCTTCCAGATTGGGCGGCGGCTGG 659  
  
QY 685 GGTCCCACTCCCTGCTCCATCCAGTGGGCTGCGACCGCGGAGTTCCTTAAAGGA 744  
Db 660 GGTCCCACTCCCTGCTCCATCCAGTGGGCTGCGACCGCGGAGTTCCTTAAAGGA 718  
  
QY 745 ATTGACTTCAGAGGAGTGGAGGCTCCAGACCCAGGCGAGTGGTGGAGGGGCTTT 804  
Db 719 ATTGACTTCAGAGGAGTGGAGGCTCCAGACCCAGGCGAGTGGTGGAGGGGCTTT 777  
  
QY 805 CCAAGAGAGAGGCTGG-----TCAAGCAGAGCGGCGGCTGTGCCCGCAGGTCCTGGAGGC 860  
Db 778 CCAAGAGAGAGGCTGGTTCAGCAGAGAGCGGCGGCTGTGCCCGCAGGTCCTGGAGGC 837  
  
QY 861 ---AGACTCGAGGCGCGAATTGTTTCTAGTTAGGCCAGGCTCTCTTCTAGTTCGCAAG 917
```

```
Db 838 AAGCTCGAGGCGCGAATTTCTAGTTAGC--AGCTCTCTCTGTCAGGTCGCAAG 895  
QY 918 GTGAACACTCATCGGAGCGCATGGGCGCTCTGAGCAACTGTGCGAGACCTTTTCAACCCC 977  
Db 896 GTGAACACTCATCGGAGCGCATGTG-CCTTTGAGCAACTGTGCGGCGCTTACCCATT 954  
QY 978 AA 979  
Db 955 AA 956  
  
RESULT 15  
BE797127  
LOCUS BE797127 874 bp mRNA linear EST 20-SEP-2000  
DEFINITION 601587116F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3941595 5',  
mRNA sequence.  
ACCESSION BE797127  
VERSION BE797127.1 GI:10218325  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 874)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCM793 row: o column: 04  
High quality sequence stop: 734.  
Location/Qualifiers  
1..874  
/organism="Homo sapiens"  
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/clone="IMAGE:3941595"  
/tissue_type="small cell carcinoma"  
/cell_line="MGC3"  
/lab_host="DHIOH (phage-resistant)"  
/clone_lib="NIH_MGC_7"  
/note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5',  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
  
ORIGIN  
  
Query Match 77.7%; Score 772.2; DB 10; Length 874;  
Best Local Similarity 96.5%; Pred. No. 1e-145;  
Matches 822; Conservative 0; Mismatches 24; Indels 6; Gaps 3;  
  
QY 24 GATGCGGACGAGGAGAGCTGCGCGCGCTGGGAGAGCGCATGAGCCGAGCTCAGG 83  
Db 1 GATGCGGACGAGGAGAGCTGCGCGCGCTGGGAGAGCGCATGAGCCGAGCTCAGG 60  
  
QY 84 CCGAGTGTACTACTTCAACCACTACATCAAGCCAGCGAGCGGCGCGCGCA 143  
Db 61 CCGAGTGTACTACTTCAACCACTACATCAAGCCAGCGAGCGGCGCGCGCA 120  
  
QY 144 CAGCAGCAGTGTGGCAAAACGGGCGAGGCGGAGCGCTGCCAGGTCGCTCGACCT 203  
Db 121 CAGCAGCAGTGTGGCAAAACGGGCGAGGCGGAGCGCTGCCAGGTCGCTCGACCT 180
```

```
QY 204 GCTGTGAAGCACAGCCAGTACAGCGGCGCCCTCTCTGGCGGACAGGAGAGATCACCCG 263
Db 182 GCTGTGAAGCACAGCCAGTACAGCGGCGCCCTCTCTGGCGGACAGGAGAGATCACCCG 240
QY 264 GACCAAGAGAGAGGCGCTGGAGCTGATCAACGGCTACATCCAGAGAGATCAAGTCGGGAGA 323
Db 241 GACCAAGAGAGAGGCGCTGGAGCTGATCAACGGCTACATCCAGAGAGATCAAGTCGGGAGA 300
QY 324 GGAGGACTTTGAGTCTCTGGGCTCAAGTTGAGGACTGCAAGCTCAGCCAAAGGCCAGGGG 383
Db 301 GGAGGACTTTGAGTCTCTGGGCTCAAGTTGAGGACTGCAAGCTCAGCCAAAGGCCAGGGG 360
QY 384 AGACTGGGTSCCTTCAGCAGAGTCAAGTCCAGAGAGCCATTTGAGAGCGCTCGTTTSC 443
Db 361 AGACTGGGTSCCTTCAGCAGAGTCAAGTCCAGAGAGCCATTTGAGAGCGCTCGTTTSC 420
QY 444 GCTCGGACCGGGGAGATGAGCGGCGCCCTGTTCACGGATTCCCGSCATCCACATCATCCT 503
Db 421 GCTCGGACCGGGGAGATGAGCGGCGCCCTGTTCACGGATTCCCGSCATCCACATCATCCT 480
QY 504 CCGCACTGAGTGAGGTGGGAGCCAGGCGCTGCGCTCGGGGACAGGGCGGCTAGG 563
Db 481 CCGCACTGAGTGAGGTGGGAGCCAGGCGCTGCGCTCGGGGACAGGGCGGCTAGG 540
QY 564 CCGGCCAGCTCCCGCTTGGCGCCAGCCAGTGGCGCGAACCCTCCACTCCCTGCCACCGTC 623
Db 541 CCGGCCAGCTCCCGCTTGGCGCCAGCCAGTGGCGCGAACCCTCCACTCCCTGCCACCGTC 600
QY 624 ACACAGATTATTTGTTCCACAAATGGCTGGGAGGGGCGCTTCCAGATTGGGGGCGCTG 683
Db 601 ACACAGATTATTTGTTCCACAAATGGCTGGGAGGGGCGCTTCCAGATTGGGGGCGCTG 660
QY 684 GGGTCCCACTCCCTGTCCATCCCAAGTTGGGGGTGCGACCGCCAGATTCTCCCTTAAGG 743
Db 661 GGGTCCCACTCCCTGTCCATCCCAAGTTGGGGGTGCGACCGCCAGATTCTCCCTTAAGG 720
QY 744 AATTGACTTCAGCAGGGGTGGAGGCTCCAGACCCAGGGCAGTGTGGTGGAGGGGTGT 803
Db 721 CATTGACTTCAGCAGGGGTGGAGGCT-CCAGAACCCAGGTGTGTGGCGAAGGGTGT 779
QY 804 TCCAAAGAGAGGCGCTGGTACAGCAGAGCCCGCCCTGTCTCCCGCCAGGTGTGGAGGCGAGA 863
Db 780 TCC-AAAGAGAGGCGCTGGAC-SCAAAGCGCCCGGGTCCCTCCNGGTGTGGAGGCGAC 834
QY 864 CTCGAGGGCGGA 875
Db 835 TCCGAGGGCCAA 846
```

Search completed: June 28, 2004, 18:01:43  
Job time : 3452.82 secs